Analyse & Kritik 02/2012 (© Lucius & Lucius, Stuttgart) S. 315–338

Bernd Lahno

Simple Games of Information Transmission*

Abstract: Communication is an inherently strategic matter. This paper introduces simple game theoretic models of information transmission to identify different forms of uncertainty which may pose a problem of trust in testimony. Strategic analysis suggests discriminating between trust in integrity, trust in competence, trust in (the will to invest) effort and trust in honesty. Whereas uncertainty about the sender's honesty or integrity may directly influence a rational receiver's readiness to rely on sender's statements, neither uncertainty about the competence of a sender nor uncertainty about his willingness to invest effort has any direct impact on rational reliance on its own. In this regard, trust in honesty and trust in integrity appear to be more basic than trust in competence or effort.

1. Introduction

At the root of social epistemology is a simple insight: most of what we know stems from communication with others. Classical epistemology has, more or less, suggested that knowledge is the result of an individual endeavor based on personal perception, memory and rational reflection. But, as a matter of fact, only a very small part of what an average person knows can actually be traced back purely to his own personal experiences and deliberations. We learn about the world from what our parents and teachers tell us, from listening to others, reading books or watching TV. A background of information transmitted by others permeates even the most simple truths that an individual may possess. In fact, none of us is capable of grounding his or her individual body of knowledge in his or her individual experiences or deliberations. That the knowledge an individual possesses is to a great extent dependent on being informed by others is even more obvious in science. Science is an ongoing collaborative project, building on the effort and achievement of many. Every new piece of scientific knowledge added is fundamentally built on the results of others. Today many results are themselves produced by cooperative work of several researchers with

* I am deeply indebted to Tobias Steinig. He continually read and criticized previous versions of this paper. Discussing things with him was invaluable help in clearing up my thought. Moreover his help in writing an English text was crucial for the readability of this paper.

different expertise, each of whom contributes only a part without being in a position to actually control what others contribute.

The emergence and development of knowledge cannot be understood by looking at isolated individuals only. At least a substantial portion of our individual knowledge is gained in an interactive process with people—intentionally or unintentionally—transmitting information, exchanging arguments, trying to convince each other, and, of course, occasionally also deceiving each other. The analysis of these interactive processes is one of the aims of social epistemology—that branch of epistemology that deals with "the relevance of social relations, roles, interest, and institutions to knowledge" (Schmitt 1994, 1).

What makes interaction special as compared to isolated individual action is that in interaction actions are connected by a curious interdependency. What people do is generally understood as being dependent on their intentions and their beliefs about the world. In interaction the relevant part of the world is essentially determined by the expected action of others. So in making his decision, a person A will typically have to reflect on the intentions and beliefs of another person B; and these beliefs of B, which are the object of A's reflection and are to be the basis of his decision, will in turn—at least partly—depend in the very same way on expectations concerning A's beliefs including the one to be formed. In short: interaction is inherently characterized by *strategic* problems.

Game theory is the most prominent and elaborated tool to analyze strategic problems of social interaction. This paper is an attempt to explore the potential of game-theoretic modeling in the analysis of very basic strategic problems in communication or, more precisely, in testifying. Its objective is fairly modest. I do not intend to find long-awaited answers to fundamental problems of social epistemology. Rather, I aim at pointing out some basic but characteristic forms of strategic problems, more or less neglected at present within social epistemology, but typically arising in information transmission. And, in doing so, I will introduce some of the basic tools of game-theoretic analysis which might be useful in understanding strategic aspects in the social processes underlying knowledge.

My analysis will focus on very simple stylized situations. The idea behind this is to isolate aspects of the social practice of testifying that have a crucial impact on the possibility and the quality of testimony. Proceeding this way, I take up a familiar debate in social epistemology, namely the debate about the determinants of reliability in testimony. Most scholars identify crucial factors determining the reliability of testimony within the testifying individual. They usually discriminate between two sorts of such determinants, between those that depend on the intentions of the testifying person and those that depend on his capabilities. The former are typically summarized as the 'honesty' or 'sincerity' of the person while the latter make up his 'competence'.[1]

Problems of uncertainty stemming from imperfect knowledge of relevant characteristics of a testifying person are normally referred to as problems of trust

[1] See e.g. Goldmann 1999, 109ff.; Fricker 1987; Audi 2003, 238. In an attempt to analyze the probabilistic structure of testifying, Friedman 1987 identified 4 different sources of failure in testifying, by further discriminating between 3 different forms of competence.

in testifying. Trust is a complex concept, differently used in different contexts. In a very wide sense trust is often understood as a(ny) positive and rational expectation concerning the actions of another person (e.g. Gambetta 1988). I argued elsewhere (Lahno 2002; 2004) that this understanding does not coincide with what we usually mean by 'genuine trust', which is a specific source of such positive expectations. In this paper I will use the term '*problem of trust*' to refer to a situation with the following characteristics:

1. a person A is uncertain about some positively valued characteristic of another person B;
2. by acting on the belief that B possesses this characteristic, A would make himself vulnerable to the actions of B, i.e., A may suffer some preventable disadvantage from relying on B.

I use the term '*trust*' to refer to a specific attitude that tends to elicit positive expectations in a situation of this sort. I will not attempt to give any further specification of this attitude here (see Lahno 2002).

Obviously a situation in which a testimony on some issue is offered will in general possess the characteristics of a problem of trust. Whereas the common discrimination of aspects of reliability in social epistemology as presented above conveys a basic distinction between trust in honesty and trust in competence, my strategic analysis of basic models of testimony will indicate additional forms of trust relevant to testimonial communication, namely trust in integrity and trust in effort.

The argument will proceed as follows. After some preliminary general remarks about strategic aspects of communication and game-theoretic analysis in social epistemology (2), I will introduce a simple probabilistic model of evidence which may also serve to represent the competence of a person (3). I, then, present a basic model of information transmission on this basis, assuming the absence of any conflict of interest (4) and discuss different problems of trust that occur in such contexts: trust in integrity (5) and trust in competence (6). Adding some cost of information to the model will lead us to identify trust in effort (7). The final model introduced here abstracts from problems of competence and information costs but introduces some conflict of interest to illustrate trust in honesty (8). I will conclude with some summarizing remarks.

2. Strategic Aspects of Information Transmission

Social epistemologists are well aware of strategic aspects in the social processes involved in the production of knowledge. The central role that the concept of trust plays—in particular in much of the debate on testimony—vividly reveals this awareness.[2] There are also a few isolated attempts to exploit game-theoretic modeling in the analysis of problems of trust occurring in the social acquisition

[2] See e.g. Hardwig 1991; Govier 1993; Fricker 1994; Baurmann 2007, and, to name a most prominent source, Coady 1992, 46f.

of knowledge. The most prominent among these is probably Michael Blais's argument against anti-reductionism about testimonial justification.

Blais (1987) asserts, first, that the essence of the trust problem in information transmission processes should be analyzed in terms of a classical Prisoners' Dilemma game. He then argues, second—drawing on Axelrod's work on cooperation (1984)—that the problem may be resolved by exclusive reference to rational considerations and to information directly available to individuals in ongoing interaction.[3] This strategy results in two problems. First, Blais's making use of the theory of iterated games and of Axelrod's evolutionary theory means that his analysis has to cope with the typical weaknesses of these theories in explaining trust.[4] Secondly and most importantly, using the PD as a model for information transmission spells trouble: the PD is a highly stylized game with simple but very specific properties. By using this game as representing the fundamental problem Blais abstracts away much of the variety of aspects actually found in situations of reliance on testimony which might well be essential in understanding the trust observed.[5] For instance, unlike most situations of knowledge transmission, the Prisoners' Dilemma displays a symmetric incentive structure with a distinctive form of conflict and a distinctive symmetry in the need for information. Moreover, while the transmission of knowledge is a dynamic process, the Prisoners' Dilemma displays a situation where individuals make their choices simultaneously.

Although important social situations with a significant problem of trust may actually display the essential structural elements of a Prisoners' Dilemma, it is highly questionable whether this also applies to the typical problems of trust in processes of information transmission. A more refined strategic analysis of testimony and communication in general seems necessary to understand the role of trust in testimony.

Strategic considerations also play a major role in Goldman's program of "Systems-Oriented Social Epistemology" (Goldman t.a.).[6] The object of the program is to analyze and evaluate different institutional arrangements or 'systems' in view of their truth generating and preserving properties. The fundamental perspective is that of methodological individualism: the systemic properties are derived from individual actions within the system. The analysis is crucially guided by the insight that what actors actually seek in their actions is not necessarily truth. They may not primarily be interested in their own knowledge and particularly not in the knowledge of others; in fact, they may be interested that the truth be hidden from others. So the aim is to characterize those arrangements that set suitable incentives so as to promote the spreading of knowledge optimally. This is an inherently strategic problem. However, the strategic analysis within the program seems to be somewhat restricted; strategic considerations,

[3] A similar argument may be found in Rescher 1998, 33ff.

[4] See Lahno 2002 for a detailed critique of the 'traditional' attempt to explain trust as a rational expectation in ongoing interaction.

[5] See Woods 1989 and Hardwig 1991 for a detailed critique from this point of view.

[6] The principal idea of the program is already evident in his groundbreaking "Knowledge in a Social World" (Goldman 1999).

as a rule, concentrate on those actors that produce and transmit information, leaving out the recipients of this information.

An illuminating example is Roger Koppl's game-theoretic analysis of 'epistemic systems' (Koppl 2005). An epistemic system consists of a set of 'senders', a set of 'receivers' and a set of 'messages', i.e. the strategies senders may choose to inform receivers. The object is to analyze which messages senders choose under different institutional constraints and to evaluate the institutional arrangements on this basis in terms of 'epistemic efficiency'. In Koppl's model receivers' utility functions are exclusively defined by the messages chosen by senders. Consequently, only the choices of senders determine the consequences of interaction. Receivers have no control over the result of interaction; they are more or less passive recipients of messages, not acting participants in the interaction. So, strategic analysis is confined to the interaction of different senders, while receivers and their utilities (or reactions to messages) are exogenous, part of 'the rules of the game'. But this entails a serious restriction of the analysis. One may easily imagine that receivers will evaluate messages in the light of their knowledge of how those messages are motivated. So there might well be strategic interdependency between messages sent and the evaluation of these messages. In Koppl's model it is impossible to account for such an interdependency. The relationship between messages and their evaluation is, in an important sense, seen as a one-way process: messages are chosen in the light of expectations about their evaluations, but the evaluation is given independently.

A possible reason for this theoretical restriction may be the specific sort of reaction characteristic of the recipient of a message. We usually think of this reaction as consisting in adopting certain beliefs. However, conventional philosophical wisdom has it that our beliefs are not under our direct control; at least they are not under our control in the way that acts are: we do not choose our beliefs in the way we choose our acts.[7] Therefore, it seems reasonable that the recipient of information is, as a recipient, not an actor and, consequently, not a participant in strategic interaction.

It seems to me that the difference between belief and act in their relation to will is somewhat overstated,[8] but I will not discuss the matter here. However this general issue may be resolved, the recipient of a message will typically act on the belief formed on the basis of the information transmitted. This very fact may be subject to strategic deliberations of the sender, which, in turn, if known to the recipient will influence his interpretation of the message received. Thus, there is a strategic interdependence between sending a message and interpreting this message.

The first person to explicitly notice this interdependence between message and interpretation was probably David Lewis in his seminal work on conventions (1969). Lewis argued that certain signaling systems—consisting of a sender

---

[7] See Williams 1973 for a classical argument in favor of this distinction. In social epistemology this distinction is also emphasized, e.g., in Webb 1993 and Goldman t.a.

[8] A solution to a dynamic game of imperfect information typically consists of a profile that assigns a strategy and a set of beliefs to each actor. The principle idea behind this is that actors choose their actions and beliefs simultaneously in solving the strategic problem.

sending a message and its receiver possibly acting on this message—have the strategic properties of a coordination problem. He introduced and analyzed simple coordination games as abstract models of such problems. The strategic interdependency of signal and interpretation is captured in these models by making the interpretation of the message inherent in the behavioral reaction to the message. I will follow Lewis's interpretation here. So, believing that p will be incorporated as acting on the assumption that p in the models I propose.

Lewis was interested in the conventional foundation of language. He was not very much concerned with epistemological issues such as truth and the transmission of truth in testifying. Thus, he concentrated in his analysis on situations in which truth is not the problem. In his models he generally assumed that there is no uncertainty—neither on the side of the sender nor on the side of the receiver—about the sender being in the possession of some bit of true information not directly available to the receiver. Moreover, he presupposed that sender and receiver by and large have a common and shared interest in the transmission of this information. Both assumptions are crucial to reducing the strategic character of the target situations to that of a simple coordination problem. In trying to extend Lewis's principal view of communication as a strategic problem to the realm of social epistemology, I will abandon these two assumptions. I will allow for some conflict of interest between sender and receiver. And, what is even more important, I will include some uncertainty as to whether the sender truly has information valuable to the receiver.

Including this sort of uncertainty requires some additional technical effort. Whereas Lewis's models are simple two-person normal form games, I will introduce dynamic games of imperfect information. Whereas Lewis could base his analysis on the classical concept of Nash-equilibrium, dynamic games of imperfect information demand some Bayesian analysis and an extended equilibrium concept. However, I will keep technical details to a minimum, concentrating instead on the guiding ideas and the principal results of the analysis.

My analysis is also related to a prominent topic in game theory and microeconomics, namely to the theory of so called signaling games.[9] Although the technical apparatus used in the analysis of signaling games is indeed identical with the one my analysis is based on and although the term 'signaling games' may suggest a general concern with messages sent and their interpretation, the theory is rather restricted with respect to transmission of general information. The core problem of the theory of signaling games is of the following form: an individual A has private information on his own characteristics which might be relevant to another individual interacting with A. The principal question, then, is: under what conditions will A be able and willing to uncover his hidden characteristics (usually by costly actions)? This core problem is too restricted to meet the general needs of social epistemology. First and probably most important, the focus is on a very special kind of information—namely information on personal characteristics—, which seems to have rather minor or secondary importance in the social production of knowledge. Second, there is always some conflict of

[9] See e.g. Gibbons 1992 for an introductory overview. An interesting recent application of the theory of signaling games is Gambetta 2009.

interest in the background: if signaling a certain characteristic is advantageous, signaling may well be advantageous regardless of whether the person actually has the characteristic or not. Finally, the signals analyzed are of a specific kind, they are typically acts but not speech acts and, thus, unlike the typical communication acts in the social process of information transmission as in testifying. Nevertheless, the underlying strategic problems are quite similar and there is much to learn from the technical details of the theory of signaling games. My project here is rather moderate in this respect, too. I will concentrate on comparably simple games and employ analyzing tools which are presumably easily comprehended.

Using game theory as a tool in analyzing strategic interaction implies a specific point of view: the perspective of methodological individualism. Within social epistemology the reach and the adequacy of methodological individualism is disputed. Personally, I have every confidence in the strengths of this methodological approach. But I do not want to argue for it here. I will just adopt this perspective as a methodological guideline. It seems to me that even a person deeply (and, in my view, falsely) convinced that social phenomena can in their essence never be adequately comprehended by reduction to individual acts should and will be interested in the possible reach of an account from an individualistic point of view.

3. Evidence and Competence

The basic problem that we are interested in can be illustrated by the following simple situation: Suppose one person, the 'recipient', has some interest in getting to know some matter of fact, and another person, the 'sender', has—with some positive probability—private evidence as to that matter. Under what conditions will the sender be willing and able to communicate his private information successfully (and truthfully) to the recipient? This may seem to be an easy question, but a closer look soon reveals that it is, in fact, quite complex and difficult. To identify at least some of the conditions we will keep things particularly simple.

We will first assume that the matter of fact in question comes in a simple Yes-No form. Something is either the case (φ) or not ($\neg\varphi$) and people are interested to know. It is further assumed that there is a prior probability known to all individuals involved that φ applies: $\mathrm{p} = \mathrm{p}(\varphi) \geq 0.5$.[10] As a convenient illustration let us assume that Robinson, the recipient, is in his base-camp in the morning considering whether to go fishing or hunting. On some days there is plenty of fish in the sea around the island and if this is so Robinson wants to go fishing. Otherwise it is best to go hunting. So Robinson has an interest in knowing whether there is fish or not.

Now, Sunday is situated on a nearby hill, where he can watch the sea and there he might gather some additional evidence on whether there is fish or not by watching the birds fly above the surface of the sea. So here is a second person

[10] If the probability is smaller than 0.5 one can simply rename events.

with access to some extra information, a (potential) 'sender'. We assume that the prior probability of fish as known to Robinson and Sunday is at least as large as the probability of no fish. So 'fish' is denoted by 'φ' and 'no fish' by '$\neg\varphi$'.

Let us have a closer look at the situation of the sender S first. We model the fact that S might have extra information by introducing an indicator I that is sufficiently related to whether φ applies or not and whose state is known to S. Again, to keep things simple we assume that the indicator has exactly two states. Thus, the indicator may be given by some proposition ψ (which might be but is not necessarily identical to φ): The indicator state is I_1 if and only if ψ is true and I_2 otherwise. Knowing the indicator state amounts to knowing whether ψ is true or not. In our exemplary illustration the indicator I would be given by 'There are birds flying above the surface of the sea'.

The correlation between the indicator states and φ is given by two conditional probabilities: If φ is the case, then there is a probability $q_1 = p(I_1|\varphi)$ that the Indicator state is I_1, while, if $\neg\varphi$ is the case, the probability of the indicator state being I_2 is $q_2 = p(I_2|\neg\varphi)$. We assume that S knows the conditional probabilities q_1 and q_2. *Figure 1* illustrates the information as given to the sender.[11]

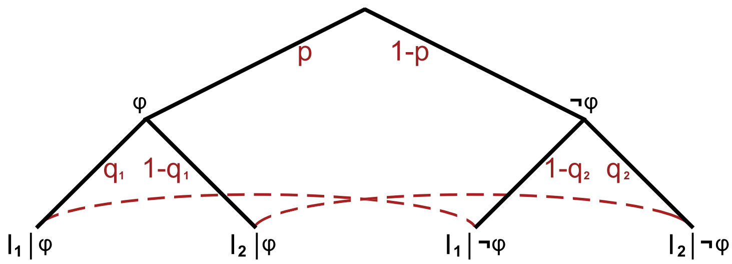

Figure 1

Unless ψ is identical to φ, realizing the state of the indicator does not amount to knowing whether φ is the case or not. This is indicated by the curved lines connecting $I_1|\varphi$ and $I_1|\neg\varphi$ and, respectively $I_2|\varphi$ and $I_2|\neg\varphi$. $I_1|\varphi$ and $I_1|\neg\varphi$ form an 'information set': the sender knows that the indicator state is I_1, but, as the indicator state might be I_1 with φ being the case as well as with $\neg\varphi$ being the case, S cannot deduce with certainty whether φ applies or not. Still knowing the indicator state gives him some information on the probability of φ and $\neg\varphi$. Using Bayes' rule we have:

$$p(\varphi|I_1) = \frac{p(I_1|\varphi)p(\varphi)}{p(I_1|\varphi)p(\varphi) + p(I_1|\neg\varphi)p(\neg\varphi)} = \frac{q_1 p}{q_1 p + (1-p)(1-q_2)}$$

[11] Friedman 1987 applies a similar argument to the situation of a recipient of some testimonial evidence who determines the probability that some proposition is true given the testimonial statement of another person (see also Goldman 1999, 109ff.). However, to give a full account of strategic problems in testimony it seems adequate to locate a problem of evidence at the position of the sender already.

$$p(\neg\varphi|I_1) = \frac{p(I_1|\neg\varphi)p(\neg\varphi)}{p(I_1|\varphi)p(\varphi) + p(I_1|\neg\varphi)p(\neg\varphi)} = \frac{(1-p)(1-q_2)}{q_1 p + (1-p)(1-q_2)}$$

Thus, knowing that the indicator state is I_1, the sender will consider φ more probable than $\neg\varphi$ iff:

$$q_1 p > (1–p)(1–q_2) \Leftrightarrow q_1 p + q_2 (1–p) > 1–p$$

Accordingly we get:

$$p(\varphi|I_2) = \frac{p(I_2|\varphi)p(\varphi)}{p(I_2|\varphi)p(\varphi) + p(I_2|\neg\varphi)p(\neg\varphi)} = \frac{(1-q_1)p}{(1-q_1)p + (1-p)q_2}$$

$$p(\neg\varphi|I_2) = \frac{p(I_2|\neg\varphi)p(\neg\varphi)}{p(I_2|\varphi)p(\varphi) + p(I_2|\neg\varphi)p(\neg\varphi)} = \frac{(1-p)q_2}{(1-q_1)p + (1-p)q_2}$$

Knowing that the indicator state is I_2, the sender will consider $\neg\varphi$ more probable than φ iff:

$$(1–p)q_2 > (1–q_1)p \Leftrightarrow q_1 p + q_2 (1–p) > p$$

Note that $q^* := q_1 p + q_2(1–p)$ denotes the probability that either φ applies and the indicator state is I_1 or $\neg\varphi$ applies and the indicator state is I_2. Thus, if the indicator is understood as indicating φ by I_1 and $\neg\varphi$ by I_2, q^* determines the overall probability that the indicator is 'right'. Again, we assume that $q^* \geq 0.5$.[12]

Using $p \geq 0.5 \geq 1 - p$, we get the following results:

- If $q^* > p$ then the indicator is decisive in the following sense: φ is more probable than $\neg\varphi$ if and only if the indicator state is I_1 and $\neg\varphi$ is more probable than φ if and only if the indicator state is I_2.
- If $q^* < p$ then φ is more probable than $\neg\varphi$ irrespective of the state of the indicator.
- If $q^* = p$ then φ is at least as probable as $\neg\varphi$ whatever the state of the indicator. Or, more precisely, φ is more probable than $\neg\varphi$ with the indicator state being I_1 and just as probable as $\neg\varphi$ with indicator state I_2.

q_1, q_2 and q^* represent the reliability of the indicator.

One of the advantages of modeling evidence by an abstract indicator is that the indicator may in principle account for all sorts of different information. In our exemplary illustration the indicator was given by 'There are birds flying above the surface of the sea'. But sender Sunday may occasionally perceive birds although there are none, he may, for instance, mistakenly perceive birds although there is only a certain formation of fog above the sea. What Sunday actually knows, then, is not whether there are birds or not but wether he perceives birds

[12] If it turns out that $q_1 p + q_2(1–p) < 0.5$, we simply rename the indicator states.

or not. This is easily captured in our model by making the indicator being determined by 'Sunday perceives birds above the surface of the sea' instead of 'There are birds flying above the surface of the sea'. Thus, the indicator may represent both internal and external determinants of the extra information as given to the sender. And, consequently, in our model q_1, q_2 and q^* may represent different aspects of reliability, external as well as internal to the sender.

Goldman (1999, 123) discriminates three different elements that determine the reliability of testimonial information: (A) the sender's competence, i.e. his "ability to detect and retain the kind of event or fact in question", (B) his opportunity "to detect the putative fact in question", and (C) his sincerity or honesty in transmitting the information. Whereas honesty refers to the transmission of information, both competence and opportunity refer to elements in the process of information generation. Competence and opportunity determine the additional information about φ as perceived by S. Although competence is in some natural sense internal to the sender while opportunity is external, both competence and opportunity are not subject to the decisions of S. They are simply given: constraints in his search for information. The indicator in our model may represent all aspects of information supply to and information processing of the sender that are not under S's direct control. Thus, within our model there is no essential difference between competence and opportunity as determinants of the evidence given to the sender. Both simply determine an indicator with certain conditional probabilities of the same general form. Both are conceived of as (externally given) constraints on information processing.

Consider a case of pure opportunity: Sunday is a perfect observer of fish flying above the surface of the sea. Whenever there are birds, he will see them. But on some days birds are miraculously absent from the island and the surrounding sea. On all other days birds are a perfect indicator of fish: there is fish if and only if birds are flying above the surface of the sea. Sunday does not know when birds are absent, but he knows the probability q that a day is such an unfortunate day without birds. In this totally unrealistic and theoretical situation the information Sunday receives depends on external factors only: on whether there is fish and on whether it is a normal or one of the unfortunate days. If I_1 is given by 'Sunday perceives birds' and I_2 by 'Sunday perceives no birds', then $q_1 = 1 - q$ and $q_2 = 1$.

Consider, in contrast, a case of pure competence: birds of a certain species, 'A-birds', are an unmistakable indicator of fish. Whenever there is fish, there will also be A-birds; and whenever there is no fish no A-birds are flying above the surface of the sea. But telling A-birds from other birds requires a lot of experience and some difficult reasoning. So Sunday sometimes thinks there are A-birds although there are none and sometimes there are A-birds but Sunday does not identify them correctly. Sunday's competence, then, can be exactly described by the conditional probabilities of the indicator states as given by 'Sunday thinks there are A-birds' and 'Sunday thinks there are no A-birds'.

Of course in any more realistic case, elements of opportunity as well as elements of competence will aggregate to the conditional probabilities of the indicator states. Nevertheless, in principle, opportunity and competence are of the same basic structure and the aggregation proceeds along the basic rules of probability theory.

4. Information Transmission: No Conflict, Complete Information

Now suppose the sender can transmit his information by using one of two signals S_1 or S_2. Sunday may, e.g., lift his arm or leave his arm unlifted and Robinson can—without any possibility of error—see what Sunday does. Under what conditions may the signals suffice to communicate what Sunday knows about fish? From the perspective of decision theory this is a matter of interest. What is the reason for Robinson's interest in the information? What interest does Sunday have in Robinson knowing about the fish? And, what do they know about each other's interests?

In the most simple case, both individuals have completely congruent interests; they are completely informed about their respective interests, about the external conditions and the opportunities and competences they have. The extensive-form game in *figure 2* represents such a simple case.

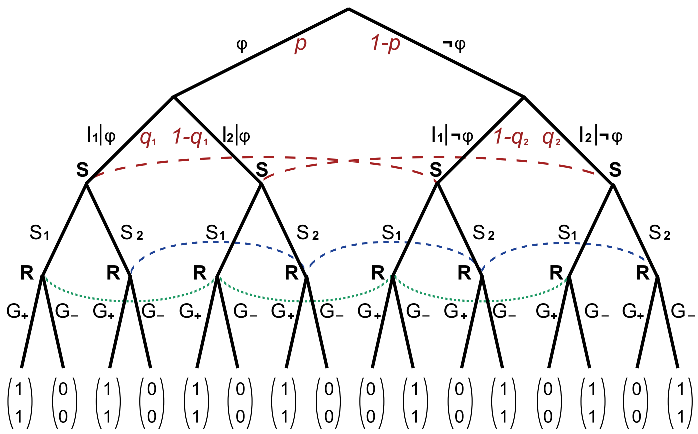

Figure 2

After receiving the signal given by the sender S the recipient R can choose between two actions G_+ and G_-. G_+ represents 'act on the assumption that φ' (Robinson goes fishing), whereas G_- represents 'act on the assumption that $\neg\varphi$' (Robinson goes hunting). R identifies the signal given, but he neither knows whether φ or $\neg\varphi$ applies nor what state of the indicator S perceived. R cannot discriminate between the knots connected by the green or blue curved lines. So he can only make his decision dependent on the signal received. Thus R has 4 different strategies:

g_+: Choose G_+ whatever signal you receive.

g_1: Choose G_+ if you receive signal S_1 and G_- if you receive signal S_2.

g_2: Choose G_- if you receive signal S_1 and G_+ if you receive signal S_2.

g_-: Choose G_- whatever signal you receive.

Following strategy g_+ (or g_-) R just acts on the assumption that φ (or $\neg\varphi$, respectively) irrespective of the signal given. If R follows g_1 he interprets S_1 as meaning φ (fish!), S_2 as meaning $\neg\varphi$ (no fish!), whereas with g_2 he acts on a reverse interpretation.

It is assumed here that R generally prefers to act on a true assumption. He values all outcomes resulting from acting on a true assumption equally: his utility in these cases is normalized to 1 as given in the upper values of the terminal knots in the graph. Similarly, he values all outcomes resulting from acting on a false assumption equally with utility 0.[13] Under these conditions, a rational R wants to act on φ if and only if he believes that φ is more probable than $\neg\varphi$. He does know the prior probability of φ, but he also knows that S may have additional information; and thus, he can estimate φ more accurately. Moreover, it is also assumed that R knows the conditional probabilities q_1 and q_2 (and, thus, also q^*).

In the simple case introduced here there is no conflict: S has an identical utility function. S wants R to act on true assumptions; and, thus, he is interested to communicate truthfully. S can make his signaling decision dependent on the extra information given to him by the indicator. Just as R, S has 4 strategies:

s_1: Choose S_1 whatever the state of the indicator.

w_1: Choose S_1 if the indicator state is I_1 and S_2 if the state is I_2.

w_2: Choose S_2 if the indicator state is I_1 and S_1 if the state is I_2.

s_2: Choose S_2 whatever state of the indicator.

If S acts on s_1 or s_2, nothing can be inferred from his choice—S does not communicate any information to R. However, if he acts on w_1 or w_2 and R knows this, then R will be able to infer what S knows about the state of the indicator; as R knows the conditional probabilities, he has the same information regarding φ at his disposal as S. Note that R may not know anything about the actual properties of the indicator—in our example: Robinson may not know that Sunday watches birds to learn about fish. All he should know is: there is such an indicator and what Sunday knows about the indicator is related to φ as described by the conditional probabilities q_1 and q_2. S's signal, then, may under certain conditions directly transfer information on φ to R, whatever the indicator is like.

In keeping with the general assumptions of game theory, we assume that the structure of the game and the rationality of the actors are common knowledge: Both know the structure of the situation as presented by the game tree; in

[13] This specific assumption of uniformity in the utility functions makes the analysis considerably easier. But it has no essential impact on the general results.

particular, they know how each of them evaluates the possible results of the interaction, and what each of them knows at each stage of the game. Both are rational and know that both are rational. They know that they know all this—they know that they know that they know all this, ... and so forth.

Under these conditions, communication is easily possible if the indicator is sufficiently reliable. Suppose $q^* > p$. In this case the indicator is decisive: φ is more probable than $\neg\varphi$ if and only if S perceives I1, and $\neg\varphi$ is more probable than φ if and only if he perceives I_2. Moreover, given the utility function of both actors both realize that it is best if R chooses G_+ if φ is more probable than $\neg\varphi$ and G_- otherwise. So both agree that R should choose G_+ if and only if S perceives I_1. Now, if S acts according to g_1 (thus signaling 'fish' after realizing I1 and signaling 'no fish' after realizing I_2) and R acts according to g_1 (thus fishing after S_1 and hunting after S_2), each maximizes his utility given the choice of the other: (w_1, g_1) is an equilibrium point of the game. An agreement to use S_1 as a signal for φ (fish) will be self-enforcing, and—if no prior communication and, thus, no explicit agreement is possible—(w_1, g_1) will possibly evolve as a convention. S will use w_1 to signal whether φ or $\neg\varphi$ is the case, and R will understand and act on the signal.

Moreover, both will get the best result they can hope for given the maximum information available to them. The probability that R will do the right thing if both act according to (w_1, g_1) is q*. So, given their utility function the expected utility of each of the two is q*, the maximum they can achieve given their information. *Table 1* shows the strategic form of our game. As the payoffs of both actors coincide, there is only one entry in each cell depicting the expected utility of both sender S and recipient R.

	g_+	g_1	g_2	g_-
s_1	p	p	1-p	1-p
w_1	p	q*	1-q*	1-p
w_2	p	1-q*	q*	1-p
s_2	p	1-p	p	1-p

Table 1

If, on the other hand, $p \geq q^*$, g_+ is weakly dominant.[14] Whatever strategy S chooses, the best thing R can do is act on the assumption that φ applies. Therefore, a rational R will ignore the signal given by S and do what is recommended by the prior probability p. A rational S will anticipate this. No communication will evolve. Signals are simply pointless.[15]

[14] If $p \geq q^*$ and $p = 1-p$, then $p = q^* = 1-q^* = 1-p = 0.5$. In this case all outcomes are valued equally. Any combination of strategies is rationally possible and no information is relevant.

[15] This result is, of course, related to Hume's famous verdict on miracles (see Hume 1975, 109ff.). The prior probability p that a miracle will not occur is very high, therefore we will disregard reports of miracles even if the reporting person is highly competent (q* large, but still smaller than p).

5. Multiple Equilibria: Trust in Integrity

There is already a fundamental problem of trust involved in the simple game as analyzed in the preceding section. But the form of trust relevant here seems to have gone widely unnoticed in social epistemology.[16]

We will concentrate on the case $q^*>p$: mutually advantageous communication is possible. Trust, then, is in fact needed because (w_1, g_1) is not the only equilibrium of the game. The game has 4 equilibria in total, namely: (s_1, g_+), (w_1, g_1), (w_2, g_2) (s_2, g_+).[17] In the first equilibrium S constantly signals S1 and thus transfers no information. Therefore it is best for R to act on his prior estimation p and hence choose G_+ whatever the signal. But if R proceeds in this way (i.e. according to strategy g_+), S cannot influence the result of the interaction; thus, s_1 is a best answer to w_1 (as any other strategy is). For the same sort of reason (s_2, g_+) is also an equilibrium. In the equilibrium (w_2, g_2) the meanings of the signals are reversed: S signals φ by S_2 and $\neg\varphi$ by S_1; R interprets this correctly by choosing G_+ on S_2 and G_- on S_1.

So there is an equilibrium selection problem or, more precisely, a coordination problem in the strict sense of David Lewis (1969). I have argued at length elsewhere (e.g. 2007; 2010) that the solution of such a problem requires mutual trust. The general line of the argument is: to coordinate on a mutual plan individuals need some mutual understanding of the common scheme of action: I will have reason to do my part in the scheme only if I have reason to expect you to do your part—and you are in the same position. So both have to trust the partner to stick to the common plan.

It seems only too natural to common sense that, once an agreement is found or a convention on the equilibrium plan has evolved, it is perfectly clear and rational that everybody should do their part in the plan. But more than rational decision making in the sense of decision theory is actually needed. Rational decision must be backed by mutual understanding of what is the right thing to do. In everyday life coordination problems we are usually very certain about the 'right' collective scheme of action; and in much the same way we are very certain that this is a collectively shared point of view. Thus, it seems to us that we are doing the only rational thing to do, overlooking the fact that the rationality of choice crucially depends on our mutual understanding and our trust in the permanence of this understanding.

So even in the simple case of information transmission discussed above there is a form of trust involved. But it is the kind of trust that any common enterprise based in some way or other on convention requires—the basic trust that is a precondition for any kind of conventional communication: *trust in integrity.*

---

[16] A possible exception is Friedman 1987. Friedman identifies 'failure of articulation' as an aspect of the reliability of a testimonial statement. Failure of articulation may include the inability to correctly apply a conventional rule.

[17] All equilibria are (part of) Bayesian equilibria and sequential equilibria. The same applies also to all equilibria introduced and discussed in the rest of the paper.

6. Pure Trust in Competence

Consider a case of pure competence. The indicator states are assumed to be exclusively determined by whether φ applies or not and by S's ability to detect this. Consequently, q_1, q_2 and q* represent the competence of S.

Now, imagine first that more than one sender is involved. If the receiver can choose among different possible senders, he may ask for advice; this is like choosing among games of the sort discussed above that differ in the parameters q_1, q_2 and q*. R will choose the one with the largest q*, because this offers him the largest expected utility. If the parameters represent the competence of the sender, we could describe this by saying that he chooses the partner in whose competence he trusts most.[18] But, as we understood the game, R actually knows the competence of his potential partners as characterized by parameters q_1 and q_2. So the more adequate description is: he chooses the partner with the best competence.

Strictly speaking, there is no trust in competence involved in the game as discussed in the preceding paragraphs. This is a simple consequence of the assumption that there is no uncertainty regarding competence. Competence is exactly reflected in q_1 and q_2, and these parameters are common knowledge among the individuals. Trust in competence can actually be a problem only if there is some uncertainty regarding competence.

As can be seen from the discussion above, coordination in information transmission for the given problem is possible as long as it is common knowledge that the competence of the sender is characterized by some $q^*>p$; in particular, the recipient does not have to know the exact size of q*. Uncertainty on the part of the recipient produces a problem only if he is uncertain whether q* is larger or smaller than p. Call the sender competent if $q^*>p$ and incompetent otherwise. So here is a simple expansion of our model to include a problem of trust in competence:

Assume that R does not know exactly how competent S is. But he does have an estimate of the probability that S is competent (i.e. competent enough to make it worthwhile to coordinate with him in the process of information transfer). That probability estimate be $r > 0$. We assume that apart from R's uncertainty about q*, the situation is exactly like the one described before: there is a shared prior probability p of φ, the individuals have identical utility functions, etc. To make things easy we further assume that S may be of two kinds: with probability r he is competent with the parameters q_1, q_2 and $q^*>p$ whereas with probability 1–r S is not competent, his competence being characterized by some parameters q'_1, q'_2 with $q'^*<p$. S does know how competent he is; R only knows the probability r of facing a competent partner (with r being common knowledge).

Using similar techniques as in the model above it is not difficult to draw a game tree representing the new situation. But this game tree will contain an ad-

[18] The situation becomes much more complicated if the receiver may receive several messages from different senders. In this case we face the problem of aggregating the different information based on different degrees of competence. Moreover, the possible strategic interaction between different senders poses a serious problem. I will not discuss this here.

ditional branch (at the beginning, the 'root') to represent the two possible states of S being competent or not. So the game tree would be rather extensive. We can do without drawing this tree as the interesting point can be easily discussed without it. We will focus on the question: under what conditions will a meaningful signal be given and understood? Or technically: are there any equilibria where the Sender S gives meaningful signals if he is competent and the recipient R understands these signals and acts on them? Of course, R cannot make his choice dependent on whether S is competent or not. This, he does not know. His strategies are, therefore, the same as in the game before.

Consider R acting according to g_1; he chooses G_+ if he receives signal S_1 (he goes fishing) and G_- after receiving S_2 (he goes hunting). If an incompetent S knows this, he will constantly choose S_1, reflecting his knowledge that this will motivate R to choose G_+, which will have the best consequences possible for both in the light of the fact that the given total information implies: φ is more probable than $\neg\varphi$. The expected utility for both is p.

If, on the other hand, S is competent, he has valuable information at his disposal. R's strategy g_1 makes it possible for S to communicate this information by choosing according to w_1. Again, both will realize their best results possible: the expected utility is $q^*>p$.

Now look at the overall situation with an incompetent S acting according to s_1, a competent S acting according to w_1 and R choosing according to g_1. Both the competent and the incompetent S maximize their expected utility given what they know about their competence, about φ and its prior probability, and given that R will act according to g_1. R does not know whether S is competent or not, but he reckons that there is a positive probability r that he is. In that case g_1 promises him the best result given the total information with an expected utility of q^*. But the same holds in the other case: if S is incompetent, the best R can expect is an expected utility of p, and this is exactly the prospect of g_1 under these conditions. The overall expected utility of R is $rq^* + (1 - r)p$. For any $r > 0$, this is the best he can expect. In sum: For all $r > 0$ (s_1, w_1, g_1) is an equilibrium with maximum expected utility for all individuals involved.[19]

This result may seem somewhat surprising: if there is the slightest chance that a partner is competent, then I should follow his advice. If this is correct, then there is no real problem of trust in competence: I may think it very improbable that you know more than I do, but if I do think that there is a tiny, e.g. a one per mille chance that you do have access to relevant information, then I should listen to you and accept your authority.

This astonishing result may in part be an artifact brought about by some of the far-reaching and unrealistic assumptions of our model. Of particular importance are possibly our assumptions about the information available to the individuals, including the common knowledge assumption usually made in game theory. To see this, notice, first, that the result relies essentially on the fact that an incompetent sender S will not hurt the recipient R. Because S knows

[19] By the same sort of reasoning, (s_2, w_2, g_2) is another equilibrium point with the same properties. There is an equilibrium selection problem as before that generates a problem of trust in integrity.

that he is incompetent, he will induce R to do what is best in the light of the fact that S has no relevant extra information: act on the prior probability. But this crucially presupposes that S actually knows that he is incompetent (and, of course, knows p). In the model this is implied by the assumption that competence is characterized by a pair of probabilities (q_1, q_2), which is known to S. In our argument we interpreted the probabilities q_1 and q_2 as objective probabilities, which characterize competence objectively and may be known by the individuals. In contrast, it is conventional in game theory to interpret probabilities in general as describing the subjective belief[20] of those individuals whose information sets are affected by the probabilities. From this point of view, p, q_1, and q_2 describe the shared subjective beliefs of S and R. The analysis of the game actually stays unchanged (although it will allow for no conclusion about the actual truth of the testimonial statement). But this perspective emphasizes the realistic possibility neglected in our model that recipient and sender may well disagree about the sender's competence.

If the actors do not agree in their probability judgments, the situation may become fairly complicated. But again an informal argument may shed some light on the problem. Imagine a sender S who believes that he is competent, although he is in fact not. Such a sender (following w_1) will give a wrong signal with some probability and, thus, the recipient R, following g_1, will act sub-optimally. If there is a considerable probability of such wrong signaling g_1 will cease to be the optimal strategy of the recipient. We will refrain from a detailed analysis here. In any case, the argument suggests that the core of a problem of trust in competence is generally not so much a problem of missing competence, but a problem of the missing capability of a person to correctly assess his own competence. Trust in competence appears as essentially trust in the self-assessment of competence of another person (if no conflict of interest is involved).

7. The Cost of Evidence

In the last two sections, we simply assumed that the extra evidence given to the sender comes without cost. But this might not be true. Sunday might have to climb a small hill to see whether there are birds or not. So to give valuable information he will have to invest some effort.

Investing effort is something under the control of an individual. Sunday can choose to incur the cost of that investment or not. We can take this into account in our model above by letting S first decide on whether to invest effort (e) or not (¬e). If he does not invest, S and R play a game as in *figure 1.* If he does invest they play a game of the same form with parameters q_{e1} instead of q_1, and q_{e2} instead of q_2. The investment is assumed to be useful in the following sense:

$$q_e{*} := q_{e1}p + q_{e2}(1\text{–}p) > p \text{ and } q_e{*} > q{*}.$$

[20] There has been an enduring debate in philosophy on whether probabilities refer to objective characteristics of some state of the world or rather to subjective belief about such states. See, e.g. Gillies 2000.

Additionally, we assume that the cost of the investment is C and applies uniformly after a decision to invest effort by S: the decision to invest lowers all of S's payoffs by C and has no impact on R's payoffs.

Consider, first, the case that the receiver can observe S's choice. Then he will know what game they are playing and the analysis above applies after S's decision. Suppose further that R and S coordinate on (s_1, g_+) if $q^* \leq p$ and on (w_1, g_1) otherwise. The expected utility of S in the game after investing effort is $q_e^* - C$. In order to decide whether to incur the cost of effort or not, S has to compare this utility with the expected utility after not investing, which is the maximum of q* and p. So if

$$q_e{}^* - max\{q^*, p\} > C$$

a rational S will invest effort in extracting information and transmit the information gained on the basis of convention (w_1, g_1); R will then interpret the signal correctly and act on it. So, there is an equilibrium of the total game with S first investing effort then reporting truthfully according to w_1 and R acting on the signal given by S according to g_1.[21] Moreover, this equilibrium is pareto-efficient and offers maximum expected utility for both actors.

The same sort of argument applies if R cannot actually observe whether S invests effort or not as long as the cost of effort is common knowledge. Then it will also be common knowledge that it pays for S to invest provided that R and S choose according to (w_1, g_1) afterwards. So, again, there is a mutually advantageous equilibrium with S first investing and afterwards signaling according to w_1 and R acting on the signal given by S according to g_1.[22]

If, on the other hand, $q_e^* - \max\{q^*, p\} < C$, the investment does not pay. S will not invest and, as C still is common knowledge, R will know this. So there will be mutual knowledge that no effort will be invested, and the analysis of the foregoing sections applies.

As before, there is no actual problem of trust in S's effort as long as there is no uncertainty about the cost of the effort. However, if there is such an uncertainty, the argument proceeds in ways similar to the case of trust in competence. Suppose R does not know what the actual costs for S's investing effort are. Assume also that R attributes some strictly positive (but possibly very small) probability to $q_e^* - \max\{q^*, p\} > C$ (costs are low). If this fact is common knowledge, then, by a similar argument as in section 5, one may show that there is an efficient and mutually advantageous equilibrium with R generally choosing according to

[21] The exact definition of this equilibrium is somewhat complicated as game theory requires to specify equilibrium strategies s_S for S and s_R for R by defining a choice for each information set in the game. A possible specification of equilibrium strategies for the case $q^* \leq p$ would be:

- s_S: Choose e first; in the remaining subgame act according to w_1 if e was chosen and according to s_1 if ¬e was chosen.
- s_R: Choose according to g_1 if e was chosen and according to g_+ if ¬e was chosen.

A possible specification for the case $q^* > p$ is:

- s_S: Choose e first; in the remaining subgame act according to w_1 whatever your first choice was.
- s_R: Choose according to g_1 irrespective of whether S invested or not.

[22] As there are fewer information sets equilibrium strategies are more easily specified. s_S could be as in the cases before while s_R would be: Choose according to g_1.

g_1, while S is investing and signaling meaningfully according to w_1 if costs are in fact low, and not investing and signaling according to s_1 if costs are high and $q^* \leq p$ or according to w_1 if costs are high and $q^* > p$.

Uncertainty about cost will not produce a specific problem of trust in effort as long as it is common knowledge that R believes in costs being low with some (possibly very low) positive probability. In this regard trust in effort is like trust in competence. However, trust in competence may pose a problem if there is disagreement about competence. There seems to be no corresponding source of a problem of trust in the case of effort. Effort seems to pose a serious problem of trust only in the context of some (other) conflict of interest or if senders with different incentives to invest effort compete.

8. Information Transmission: Conflict, But No Problem of Competence, Effort or Opportunity

Perhaps the most natural problem of trust in testimony arises if a conflict of interest is possible. In the following I will discuss a simple example of partial conflict.

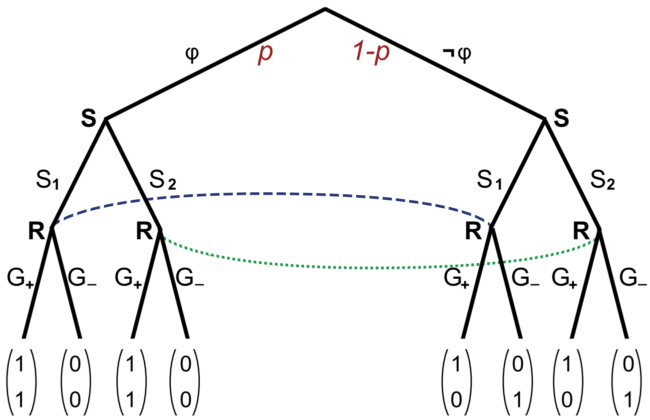

Figure 3

We begin our discussion with a closer look at conflict in testimony if that conflict is perfectly transparent to the participants of communication. The form of conflict we exemplarily address is very simple: the sender has an interest in the recipient acting on the assumption that φ applies. To illustrate this in our example, one may assume that S(unday) generally wants R(obinson) to go fishing because he is not able to fish himself and does not want to share the hunting ground. This conflict is depicted by the simple assumption that S's utility is 1 in all cases in which R chooses G_+ and 0 otherwise. To abstract from other possible problems of information transmission, we assume that there is no problem of competence or opportunity and no cost of obtaining or transmitting information involved. So C is assumed to be 0; q_1, q_2 and q^* are all assumed to be one.

If all this is common knowledge among the two participants in communication, the situation is adequately represented by the game in *figure 3*.

As the parameters q_1 and q_2 are assumed to be 1, S is perfectly informed on whether φ is the case or $\neg\varphi$ and R knows this. $I_1|\varphi$ is identical with φ and, respectively, $I_2|\neg\varphi$ identical with $\neg\varphi$. Both actors have the same strategies at their disposal as in the game in *section 3*. To give an account that captures the case that S prefers R to act on the more probable state of the world as well as the case that S prefers R to act on the less probable state, we now abandon the assumption that $p > 0.5$; p may be any positive real number smaller than 1. *Table 2* shows the normal form of the game in *figure 2* with the first number in each cell referring to the expected utility of S, and the second number, accordingly, to the expected utility of R.

If $p > 0.5$, the strategy combinations (s_1, g_+), (s_1, g_1), (s_2, g_+) and (s_2, g_2), each with the payoff combination (1 | p), are the only equilibria of the game. In equilibrium, a rational S will never transmit any information, and a rational respondent will always act on φ, i.e. he will always do what is best given his prior belief no matter what signal he receives from S. No information is transmitted.

	g_+	g_1	g_2	g_-
s_1	1 | p	1 | p	0 | 1-p	0 | 1-p
w_1	1 | p	p | 1	1-p | 0	0 | 1-p
w_2	1 | p	1-p | 0	p | 1	0 | 1-p
s_2	1 | p	0 | 1-p	1 | p	0 | 1-p

Table 2

If $p < 0.5$, there are only two equilibria, namely (s_1, g_-) and (s_2, g_-), each with the payoff combination (0 | 1-p). Again, no information is transmitted in equilibrium. R will always act according to his prior belief about $\neg\varphi$. And S will give his signal independently of his information about φ.

If $p = 0.5$, all six strategy combinations above are equilibria, and, as in both cases before, no information is transmitted in equilibrium.

The case of $p < 0.5$ is particularly interesting as the sender in this case gets his worst outcome in equilibrium. In fact, S and R share an interest in successful communication: the strategy combinations (w_1, g_1) and (w_2, g_2) offer a payoff, which is strictly higher for each of them than that offered by the equilibria. We have a typical dilemma structure here. Both would gain from meaningful and truthful communication, but the strategic character of the situation does not allow for such a solution: if R chooses according to g_1, S will do better by choosing s_1 instead of w_1. Thus, (w_1, g_1) cannot be a stable behavioral pattern, and a similar argument applies to (w_2, g_2).

The analysis above draws quite a plausible picture of the consequences of a conflict of the given sort. If I know that a person has an interest in me changing my belief, I will disregard his testimony as void. In exactly the same way, the speaker's testimony is useless to me if I know that he has an interest in endorsing my belief. In any case the other person will not be able to transmit any meaningful information about the case at hand to me by trying to signal what he wants me to believe—and this is what he will try to do as a rational individual.

But what happens if the recipient believes that there is some probability r that the sender S is 'honest' in the sense that with probability r S has utility values identical to R (as in the no-conflict case above), while with probability 1–r S is 'dishonest' with actual payoffs as given in *figure 2.* S is assumed to be informed on R's beliefs in this respect. Again, we may represent this sort of situation by introducing a first move by nature 'deciding' whether—with probability r—the no-conflict game or—with probability 1–r—the conflict game is to be played by S and R. S can observe nature's move (he knows his utility function), while R cannot, he does not know what kind of person S is beyond his estimate that S is honest with probability r. Will meaningful communication be possible in such a case?

	g_+	g_1	g_2	g_-
(w_1, s_1)	p	r + (1–r)p	(1–r)(1–p)	1–p

Table 3

Suppose R acts according to g_1. Then S's optimal answer is w_1 if S is honest, and s_1 if he is dishonest. This will give R an expected utility of $r + (1-r)p$. *Table 3* shows the payoffs which R can expect for his different strategies given S acts according to w_1 if he is honest, and according to s_1 if he is dishonest:

Notice that it is generally true that $p \leq r + (1-r)p$ and $(1-r)(1-p) \leq 1-p$. Therefore g_1 is a best answer to (w_1, s_1) if and only if $r + (1-r)p \geq 1-p$ or:

$$r \geq 1 - \frac{p}{1-p} \qquad (*)$$

Thus, (w_1, s_1, g_1) is an equilibrium iff (*) applies. This equilibrium is, in fact, pareto efficient, and S (whether honest or dishonest) will achieve his best outcome in equilibrium. By similar reasoning (w_2, s_2, g_2) is a pareto efficient equilibrium iff (*) applies. Depending on the parameters r and p the game possesses other equilibria.[23] However, in all of these other equilibria the recipient R does not obtain any relevant information from the interaction; his equilibrium strategy demands to act on his prior belief independently of the signal received (i.e. g_+ if $p>0.5$ and g_- if $p<0.5$, g_+ or g_- if $p=0.5$). Moreover, all these equilibria are strictly pareto-inferior to the (w_1, s_1, g_1) and (w_2, s_2, g_2) conventions.

Notice that in the coordination equilibria (w_1, s_1, g_1) and (w_2, s_2, g_2) R does in fact learn something from the signal given. If, e.g., the convention is (w_1, s_1, g_1) and R receives the signal S_2, then he may infer that his partner is honest (a dishonest S exclusively sends the signal S_1 in equilibrium). Therefore, he knows

[23] The number of these equilibria depends on the parameters with minimum 4 and maximum 10 if p is very close to 1.

that $\neg\varphi$ applies; he will update his prior p to p'=0. If, on the other hand, he receives signal S_1, he will update his assessment of the probability that φ, using Bayes' rule:

$$p(\varphi \mid S_1) = \frac{p(S_1 \mid \varphi) \cdot p(\varphi)}{p(S_1 \mid \varphi) \cdot p(\varphi) + p(S_1 \mid \neg\varphi) \cdot p(\neg\varphi)} = \frac{1 \cdot p}{1 \cdot p + (1-r)(1-p)} = \frac{1}{1-r(1-p)} \cdot p > p$$

Put less formally: he will assign φ a higher probability than before.

The results are just as we would expect: if there is sufficient trust that the sender is honest (and, of course, provided that no other problem of trust occurs), then testimony is useful and does, in fact, transmit information. Note that condition (*) is always valid for p>0.5. Thus, if all dishonest S want R to act on his prior belief, then the slightest possibility of facing an honest S is a good reason for R to rely on S's testimony. This suggests the quite plausible statement that it is quite easy to find people believing you if you are testifying what most people believe anyway, while you have to prove some credibility if you want to make people change their minds.

9. Conclusion

Note that in all models introduced here individuals are assumed to be already endowed with some sort of knowledge. Not only do they know something about the question at hand (as indicated by the parameter p and the indicator variables), they also know something about each other (as indicated, e.g., in the parameters q and r). In fact, it seems impossible to provide an adequate model of a situation of testimony, however simple, without such assumptions. In other words, it seems impossible to think of testifying without assuming some such knowledge. Although this result does not resolve the problem of reductionism in social epistemology, it seems at least relevant. It endorses an argument put forward by Robert Audi (2003, 137ff.): testimony is in an important sense a secondary source of knowledge. It can justify belief, but it cannot generate knowledge on its own without some anchor in already existing knowledge.

In this paper I introduced simple models of information transmission, which all display some sort of uncertainty due to imperfect knowledge. The analysis of these models allowed us to discriminate between some uncertainties about the characteristics of a testifying person from a strategic point of view: uncertainty about the integrity, the competence, the will to invest effort, and the honesty of that person. These uncertainties are related to different forms of trust that may play a role in testifying. Strategic analysis suggests that trust in honesty and trust in integrity are more fundamental than trust in effort or competence. Uncertainty about the will to invest effort seems to pose a serious problem only if combined with some other uncertainty about the intentions and the beliefs of the testifying person. Something very similar applies to uncertainty about competence. If a conflict in interest can be excluded, a real problem of trust in competence seems to arise only if there is uncertainty about the self-assessment of competence by the testifying person.

In trying to isolate specific forms of uncertainty in information transmission, I abstracted from many interesting properties that a situation in which information is transmitted by way of testimony may actually have. The resulting highly idealized models I introduced are still fairly complex and their analysis is by no means trivial. However, they represent only a limited selection of strategic problems in communication. There are others which might also be interesting, for instance the impact of the specific status of an expert on the credibility of expert reports, which seems to be a seriously strategic matter. Whether game theory and strategic analysis can shed some light on problems like these remains to be shown by further effort.

Bibliography

Audi, R. (2003), *Epistemology: A Contemporary Introduction to the Theory of Knowledge*, 2nd ed., New York

Axelrod, R. (1984), *The Evolution of Cooperation*, New York

Baurmann, M. (2007), Rational Fundamentalism? An Explanatory Model of Fundamentalist Beliefs, in: *Episteme* 4, 150–166

Blais, M. J. (1987), Epistemic TIT FOR TAT, in: *The Journal of Philosophy* LXXXIV, 363–375

Cody, C. A. J. (1992), *Testimony*, Oxford

Fricker, E. (1987), The Epistemology of Testimony, in: *Proceedings of the Aristotelian Society*, Supplementary Volumes 61, 57–83

— (1994), Against Gullibility, in: Matilal, B. K./A. Chakrabarti (eds.), *Knowing from Words*, Dordrecht, 125–161

Friedman, R. (1987), Route Analysis of Credibility and Hearsay, in: *Yale Law Journal* 96, 667–742

Gambetta, D. (1988), Can We Trust Trust, in: *Trust: Making and Breaking Cooperative Relations*, New York–Oxford, 213–238

— (2009), *Codes of the Underworld: How Criminals Communicate*, Princeton

Gibbons, R. (1992), *A Primer in Game Theory*, Hemel Hempstead

Gillies, D. (2000), *Philosophical Theories of Probability*, London–New York

Goldman, A. I. (1999), *Knowledge in a Social World*, Oxford

— (t.a.), Systems-Oriented Social Epistemology, to appear in: Szabo Gendler, T./J. Hawthorne (eds), *Oxford Studies in Epistemology. Vol. 3*, Oxford; draft available online at: http://fas-philosophy.rutgers.edu/goldman/Systems-Oriented%20 Social%20Epistemology.pdf [retrieved 8/25/10]

Govier, T. (1993), Trust and Testimony: Nine Arguments on Testimonial Knowledge, in: *International Journal of Moral and Social Studies* 8, 21–39

Hardwig, J. (1991), The Role of Trust in Knowledge, in: *The Journal of Philosophy* LXXXVIII, 639–708

Hume, D. (1975), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3rd. ed. with text revised and notes by P. H. Nidditch, Oxford

Koppl, R. (2005), Epistemic Systems, in: *Episteme* 2, 91–106

Lahno, B. (2002), *Der Begriff des Vertrauens*, Paderborn

— (2004), Three Aspects of Trust, in: *Analyse & Kritik* 26(1), 30–47

— (2007), Rational Choice and Rule Following Behavior, in: *Rationality and Society* 19, 425–450

— (2010), Individuelle Rationalität und kollektive Perspektive, in: Iorio, M./R. Reisenzein (eds.), *Regel Norm, Gesetz*, Frankfurt, 203-230

Lewis, D. (1969), *Convention. A Philosophical Study*, Harvard

Rescher, N. (1989), *Cognitive Economy*, Pittsburgh

Schmitt, F. F. (1994), Socializing Epistemology: An Introduction through Two Sample Issues, in: Schmitt, F. F. (ed.), *Socializing Epistemology: The Social Dimensions of Knowledge*, London, 1–27

Webb, M. O. (1993), Why I Know as Much as You: A Reply to Hardwig, in: *The Journal of Philosophy* LXXXX, 260–270

Williams, B. (1973), Deciding to Believe, in: *Problems of the Self*, New York–Cambridge, 136–151

Woods, J. (1989), The Maladroitness of Epistemic TIT FOR TAT, in: *The Journal of Philosophy* LXXXVI, 324–331