Stories and Assessments of Plausibility

RICKY LAM*

Department of Economics, Northwestern University, Evanston, U.S.A.

May 2001

1 Introduction

Perhaps the best known model of people's beliefs under uncertainty is that of subjective or personalistic probabilities. A decision theoretic justification for this model is provided by the appeal of Savage's (1954) axioms. In his theory, the objects of choice are *acts*, functions from a set of possible states to a set of consequences. The agent is assumed to possess preferences—formally a binary relation—over the set of acts. Savage shows that the agent's preference relation satisfies certain normative axioms if and only if she behaves as an expected utility maximizer.

"Traditional" economics does not claim that someone who satisfies Savage's axioms necessarily possesses a subjective prior and uses the maximization of expected utility as a choice procedure. Rather, the claim is that she behaves *as if* she follows such a procedure. The problem with this position, however, is that it is difficult to find plausible alternative procedures that lead to behavior consistent with the axioms. (See Rubinstein 1998.) In light of this, it is important to examine the realism of the expected-utility hypothesis—not as a representation of preferences—but as a descriptive procedure of decision-making.

Even if one were to accept subjective probabilities as a useful description of how agents view the world, questions remain. Where do subjective probabilities come from? How does the decision maker "construct" them? How is the support to subjective probabilities determined?

*Contact: rickylam@northwestern.edu

An example may help focus ideas. Consider a juror in a trial attempting to understand, or "construct beliefs", about the case in question. Here, the subjective-probability hypothesis would require the juror to form a rather complicated probability measure over states of the world—descriptions of the defendant's guilt or innocence, and all the evidence that could be presented. To proceed, she may begin by ruling out scenarios that are "incoherent" or that "do not make sense". I will refer to the remaining coherent scenarios, or coherent states of the world, as *stories*. Having established a set of stories, the juror would then have to assign some degree of confidence, or plausibility, to each possible story. For an agent who reasons with subjective probabilities, the set of stories may be loosely viewed as something akin to the support of her beliefs.

2 Stories

In a series of experiments, Pennington and Hastie (1986, 1988, 1990, 1992 and 1993) find that jurors in a trial do not seem to form probability distributions over states of the world. Nor do they incorporate information by Bayesian conditioning. Instead, they construct "cognitive representations of the evidence in the form of stories"—scenarios describing "what happened" during the events in question. Stories facilitate evidence comprehension and are constructed by jurors even though evidence is often presented out of temporal and causal order. Moreover, Pennington and Hastie find that jurors base their verdicts on the story representation of information, rather than on the raw evidence.

In fact, we often speak of constructing "stories" to facilitate learning in other real-world situations. A doctor may form a "story" to account for her patient's symptoms. A teacher may use a "story" to interpret a student's performance. A central banker may have a "story" about the current state of the economy. Lam (2001) develops a highly stylized model of how people construct stories.[1] The procedure described can be interpreted as the first step in forming subjective assessments of plausibility. The theory

[1]The mathematics of the model is a modification of the Hopfield (1982) network which has been used to solve combinatorial optimization and pattern recognition problems. The "neurons" and "synaptic connections" in our model take on different interpretations compared with these applications. The assumptions which generate many of the results are based on the interpretation we give to the mathematics.

is descriptive rather than normative.

The foundation of the model is a finite set of statements, $\mathcal{P} = \{x, ..., y\}$, that an agent has in her mind. In the trial example, statements may include: "the defendant is guilty". In order to facilitate reasoning, the agent assigns *attitudes* toward statements; these are grouped into a vector $\boldsymbol{a}$ in $\{0,1\}^{\#\mathcal{P}}$. Elements in this vector are indexed by $\mathcal{P}$. $\mathbf{a}(x) = 1$ signifies that statement x is *accepted*; $\mathbf{a}(x) = 0$ refers to a statement which is *not accepted*. Acceptance should not be interpreted as knowledge, or even as belief with probability one. Rather, the agent accepts a statement if the preponderance of evidence suggests that it is true. This interpretation will become clearer when the updating of attitudes is described.

Statements are related to one another through *reasons* which are modelled as links in two directed networks. The idea is that the attitude toward a statement can provide some justification for accepting, or not accepting, another statement. One network provides reasons from accepted statements; the other from non-accepted statements. These networks are represented by two $\#\mathcal{P} \times \#\mathcal{P}$ matrices, $\mathbf{R}$ and $\mathbf{Q}$. Elements in $\mathbf{R}$ and $\mathbf{Q}$ lie in the interval $[-1, 1]$. Cells in the matrices are indexed by statements in $\mathcal{P}$. For example, $\mathbf{R}(x, y)$ refers to the cell in row x and column y.

$\mathbf{R}(x, y) > 0$ means that the agent views the acceptance of statement y as a reason for the acceptance of x. Reasons can also inhibit acceptance: $\mathbf{R}(x, y) < 0$. The magnitude of the value attached to a link represents the strength of the reason. $\mathbf{Q}$ is defined analogously for reasons from non-accepted statements. For example, $\mathbf{Q}(x, y) > 0$ implies that the non-acceptance of y is a reason to accept x.

Shafir, Simonson and Tversky (1993) present experimental evidence which confirm that people seek reasons to justify their decisions, to themselves, and to others. The premise here is that reasons matter, not only to choices, but to attitudes.

There exists a vector ω in $\{0,1\}^{\#\mathcal{P}}$ that determines whether each statement is objectively true (1) or false (0). The task of learning involves deducing this pattern of "0"s and "1"s over the set $\mathcal{P}$. Think of ω as the binary representation of the true *state of the world*. Given ω, a signal—denoted by $\mathcal{S}$—is a subset of the statements in $\mathcal{P}$. The interpretation is that the truth values of sentences in $\mathcal{S}$ are revealed to the agent. For example, a witness's testimony may reveal to our juror that "the defendant has an alibi".

Because of reasons, a signal will typically trigger a sequence of inferences.

Continuing with the example, the juror's acceptance that "the defendant has an alibi" may lead her to accept that "the accused did not commit the crime himself". This may in turn give her some reason to revise her attitude toward the sentence "there is an accomplice to the crime", and toward the statement "the defendant is innocent", and so on.

Formally, a time subscript t is introduced in order to explicitly capture the inferential process which occurs in the mind of the agent. For example, $\mathbf{a}_t$ are the attitudes after the signal has been processed for t "periods". One more piece of notation is needed. For any subset of statements $\mathcal{E}$ in $\mathcal{P}$, and any vector $\mathbf{u}$ in $\mathbb{R}^{\#\mathcal{P}}$, let $\mathbf{u}^{\mathcal{E}}$ in $\mathbb{R}^{\#\mathcal{E}}$ denote the projection of $\mathbf{u}$ onto the coordinates representing statements in $\mathcal{E}$. For example, $\mathbf{a}^{\mathcal{S}}$ is a vector in $\{0,1\}^{\#\mathcal{S}}$ that contains attitudes toward sentences in the set $\mathcal{S}$.

Upon observing the signal $\mathcal{S}$, the agent provisionally assigns attitudes. The only requirement is that she assigns the correct attitudes toward statements whose truth values have been revealed:

$$\mathbf{a}_0^{\mathcal{S}} = \omega^{\mathcal{S}} \tag{1}$$

No requirement is made on the remaining set of attitudes, $\mathbf{a}_0^{\mathcal{P}-\mathcal{S}}$. Each period t, the attitude toward one sentence in $(\mathcal{P}-\mathcal{S})$, say x, is revised. The attitude assigned to x depends on the cumulative reason toward x, which in turn depends on the current vector of attitudes. If a preponderance of reasons suggests that the statement is true, then it remains accepted, or is revised to be accepted:

$$\mathbf{a}_{t+1}^{\mathcal{P}-\mathcal{S}}(x) = \begin{cases} 1 & \text{if } \mathbf{R}(x,.)\,\mathbf{a}_t + \mathbf{Q}(x,.)\,(\mathbf{1}-\mathbf{a}_t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Here, $\mathbf{1}$ denotes a $(\#\mathcal{P} \times 1)$ vector of ones and $\mathbf{R}(x,.)$ and $\mathbf{Q}(x,.)$ denote the xth rows of the reason matrices. The assumption of additivity in reasons is a simple way to capture deliberation in the mind of the agent. She contemplates and weighs the reasons for, and against, the acceptance of the statement; the sum in equation (2) is the result of looking for where the balance lies. Notice that the sequence of vectors that result from this procedure depends on the order (through statements in $\mathcal{P}-\mathcal{S}$) in which the agent revises attitudes.

In this framework, a story is a coherent pattern of attitudes, with the attitude toward each statement being supported and built on the attitudes toward other statements.[2] Define a story, therefore, to be a vector of attitudes, $\mathbf{a}$, satisfying:

$$\mathbf{a}^{\mathcal{S}} = \omega^{\mathcal{S}} \quad (3)$$
$$\mathbf{a}^{\mathcal{P}-\mathcal{S}} = \left(\operatorname{sgn}\left[\mathbf{Ra} + \mathbf{Q}\left(\mathbf{1} - \mathbf{a}\right)\right]\right)^{\mathcal{P}-\mathcal{S}} \quad (4)$$

Lam (2001) discusses some of the properties of this model. Among them, confusion is possible. This occurs when the agent's inferences lead her to cycle endlessly between patterns of attitudes, unable to construct a story; the juror cannot make sense of the evidence. It is shown that a contrapositive requirement[3] on the reason networks, together with irreflexivity, ensures that confusion does not occur. (Of course, the agent can still be wrong.) Depending on the order of their inferences, even agents who possess the same reasons can come to different interpretations of the same evidence: typically multiple stories can account for the evidence. And even weak signals—formally $\mathcal{S} = \{.., x, ..\}$ such that $\mathbf{R}(y,x), \mathbf{Q}(y,x) \approx 0$ for all y in $\mathcal{P} - \mathcal{S}$—can sometimes trigger a long sequence of inferences to a very different story: stories can "collapse". Lam (2001) also extend the model to account for multiple signals. With this extension, it is shown that the order of signals can affect the story that is formed.

3 Stories and Probabilities

In the current paper, we abstract from the dynamics described in Lam (2001). We focus instead on the static conditions for a story—equations (3) and (4).

[2]The premise that people like to hold coherent views of the world is supported by research in psychology. A very influential approach to explaining this human tendency was developed by Festinger (1957), who proposed that any perceived inconsistency among various aspects of beliefs, emotions, memory, and behavior, causes an unpleasant state that he termed *cognitive dissonance*, which people try to reduce whenever possible.

[3] An agent's reasons, $\mathbf{R}$ and $\mathbf{Q}$, are said to satisfy the contrapositive condition if the following hold for all $x, y \in \mathcal{P}$,

$$\begin{aligned} \mathbf{R}(x,y) &< 0 \text{ implies } \mathbf{R}(x,y) = \mathbf{R}(y,x) \\ \mathbf{Q}(x,y) &> 0 \text{ implies } \mathbf{Q}(x,y) = \mathbf{Q}(y,x) \\ \mathbf{R}(x,y) &> 0 \text{ or } \mathbf{Q}(y,x) < 0 \text{ implies } \mathbf{R}(x,y) = -\mathbf{Q}(y,x) \end{aligned}$$

We assume that the agent is aware of all stories. This is in contrast to Lam (2001), where the interpretation is that the agent reasons to one story; multiple stories never enter her mind. In real-world situations, of course, people are often aware of multiple stories. A juror may, for example, be presented with two possible stories, one from each of the opposing attorneys. Alternatively, she may undertaken more than one sequence of reasoning and come up with multiple coherent scenarios to account for the evidence. It is precisely when agents possess more than one story that an assessment of the relative plausibility of each is necessary.

To think about what the Story model implies about perceptions of plausibility, we begin by contrasting the notion of "coherence" embodied in equations (3) and (4), with that embodied in the standard probabilistic model. For simplicity, consider an agent who contemplates only two statements, $\mathcal{P} = \{x, y\}$. The true state of the world lies in the following product space: $\{0,1\} \times \{0,1\}$.

The probabilistic model assumes that all information which is relevant for decision making can be captured by probability measures. In particular, five measures may be involved:

π_x	:	Marginal measure over the truth value of x
π_y	:	Marginal measure over the truth value of y
π	:	Joint measure over the state space $\{0,1\}^2$
$\pi_{y\|\omega(x)}$	:	Measure over y, conditioned on a truth value for x
$\pi_{x\|\omega(y)}$	:	Measure over x, conditioned on a truth value for y

Bayesian beliefs can be represented in three equivalent ways: (i) by the joint measure π; (ii) by the marginal π_x and the conditional measures $\{\pi_{y|\omega(x)}\}_{\omega(x)\in\{0,1\}}$ and (iii) by the distribution π_y and the conditional measures $\{\pi_{x|\omega(y)}\}_{\omega(y)\in\{0,1\}}$. Bayesian beliefs are "coherent" in the sense that Bayes's rule allows us to move among these representations. Another way of seeing this point is via the Law of Total Probability:

$$\begin{aligned} \begin{bmatrix} \pi_x(1) \\ \pi_y(1) \end{bmatrix} &= \begin{bmatrix} 0 & \pi_{x|\omega(y)=1}(1) \\ \pi_{y|\omega(x)=1}(1) & 0 \end{bmatrix} \begin{bmatrix} \pi_x(1) \\ \pi_y(1) \end{bmatrix} \\ &+ \begin{bmatrix} 0 & \pi_{x|\omega(y)=0}(1) \\ \pi_{y|\omega(x)=0}(1) & 0 \end{bmatrix} \begin{bmatrix} 1-\pi_x(1) \\ 1-\pi_y(1) \end{bmatrix} \end{aligned} \tag{5}$$

or

$$\mathbf{p} = \mathbf{M}\mathbf{p} + \mathbf{N}(\mathbf{1} - \mathbf{p}) \tag{6}$$

where $\mathbf{p} = \left[\ \pi_x(1) \quad \pi_y(1)\ \right]'$ and the matrices M and N contain the appropriate conditional probabilities. For given matrices M and N, equation (6) can be thought of as a fixed-point condition on the marginal measures π_x and π_y. The similarity with equation (4) is obvious.

From these equations, we see at least two important differences between stories and probabilities. First, probabilities allow for degrees of belief; attitudes do not. When constructing stories, the agent recognizes that the true state of the world is binary and attempts to find a coherent view of it. A Bayesian cares about whether probabilities assigned to subsets of the state space are coherent. Second, for given generic conditional probabilities, the fixed point in equation (6) is unique. In contrast, we have said that multiple stories are possible.

As we pointed out in the introduction, one way toward a reconciliation of the two models is to view them as operating at different stages of the processing of information. In the first stage, agents eliminate states that are incoherent; this leaves a set of stories—typically substantially smaller than the original set of states. The agent then assesses which story is most plausible. These confidence assessments may take the form of probabilities. If so, one can view the Story model as providing the support for an agent's subjective probability measure.

There is one important qualification. A state that is not a story—and so not in the support of the agent's probabilistic beliefs—may become coherent after a subsequent signal.[4] Bayes's rule, however, does not allow one to update from zero to some strictly positive probability. This is particularly problematic when the set of stories after some signal have no elements in common with the set of stories prior to the signal. A procedural theory of how people make subjective confidence assessments becomes crucial.

[4]This is obviously related to the idea of unforseen contingencies. For this to be an interesting model of unforseen contingencies, however, it is necessary for the agent to forsee the possibility that the support of her beliefs might shift to incorporate new states. See Dekel, Lipman and Rustichini (1998).

4 Assessments of Plausibility

If assessments of plausibility do not derive from conditions that ensure coherency between different probabilities, but rather from the belief that the true state of the world is coherent, what can the idea of a story tell us about how agents make judgements of plausibility? Obviously, states of the world which are stories are much more plausible than states that are not. More generally, there are a number of other aspects to plausibility which seem natural in the context of the model.

Uniqueness

If, in the process of reasoning, the agent can only construct one coherent scenario, then her confidence in this story will be high. This claim finds support in Baltzer and Pennington's (1995) experimental studies.

To induce confidence, only the *perception* of uniqueness is required. This leads to the possibility of *over*confidence. If the agent perceives only one story, when in fact many exists, she will be more confident in her story than is objectively warranted. When the number of statements, $\#\mathcal{P}$, is large, and when the network of reasons is complicated, it is very easy for the agent to not "see" all the stories that exist. It is also possible that the agent begins with a provisional set of attitudes from which she cannot reason to all stories.[5] Even if all stories are reachable, finding all of them can be very costly, both in terms of effort and time.

Strong Reasons

If a story is supported by strong reasons, both in an average sense, and also in the sense that the strength of the weakest link in a story is large, then the story is more plausible to the agent. The strength of the weakest link determines how "fragile" a story is.

Explanatory Power

Recall the definition of a story in equations (2.3) and (2.4). If a subset of the signal, $\mathcal{S}' \subset \mathcal{S}$, generates the same story as the original signal, $\mathcal{S}$, then it can "explain" the evidence $\mathcal{S} - \mathcal{S}'$. If a story is associated with a set $\mathcal{S}'$ that is small, then it is more plausible.

---

[5]Simple examples can be generated using reason networks which are disconnected.

Likelihood Assessments of Events

If all stories assign the same attitude toward a statement x, then one would expect that the agent has high confidence regarding this attitude. If, in addition to having the same attitude among all stories, a statement's attitude is robust to signals, then its attitude is even more plausible.

Simplicity

One aspect of simplicity is the number of sentences that a story incorporates.

Local Uniqueness

It turns out that states of the world which are local to each other cannot both be very plausible. This is because a state which differ from a story only in the attitude toward one statement cannot be story.

Proposition (Local Uniqueness). *Assume that the reason matrices* $\mathbf{R}$ *and* $\mathbf{Q}$ *satisfy irreflexivity:*

$$\mathbf{R}(x,x) = 0 \tag{7}$$

$$\mathbf{Q}(x,x) = 0 \tag{8}$$

Then for any signal $\mathcal{S}$ *which is not completely informative* ($\#\mathcal{S} < \#\mathcal{P}$)*, two vectors of attitudes which are both stories must differ in the attitude toward at least two statements.*

Proof. Let $\mathbf{a} \in \{0,1\}^{\#\mathcal{P}}$ and $\mathbf{b} \in \{0,1\}^{\#\mathcal{P}}$ be two stories for a given set of reasons $\mathbf{R}$ and $\mathbf{Q}$, and a given signal $\mathcal{S}$. Obviously, $\mathbf{a}^{\mathcal{S}} = \mathbf{b}^{\mathcal{S}}$. Suppose, contra-hypothesis, that they only differ over one statement in $\mathcal{P} - \mathcal{S}$, say x. For this statement,

$$\begin{aligned} \mathbf{a}(x) &= \operatorname{sgn}\left[\sum_{y\in\mathcal{P}} \mathbf{R}(x,y)\,\mathbf{a}(y) + \sum_{y\in\mathcal{P}} \mathbf{Q}(x,y)\,(1-\mathbf{a}(y))\right] \\ &= \operatorname{sgn}\left[\sum_{y\in\mathcal{P}} \mathbf{R}(x,y)\,\mathbf{b}(y) + \sum_{y\in\mathcal{P}} \mathbf{Q}(x,y)\,(1-\mathbf{b}(y))\right] \\ &= \mathbf{b}(x) \end{aligned} \tag{9}$$

which contradicts the assumption that $\mathbf{a}(x) \neq \mathbf{b}(x)$. The first equality follows from the definition of a story; the second equality holds because of irreflexivity ($\mathbf{R}(y,y) = 0$ and $\mathbf{Q}(y,y) = 0$ for all $y \in \mathcal{P}$), and because, by assumption, $\mathbf{a}(y) = \mathbf{b}(y)$ for all $y \in \mathcal{P}\backslash\{x\}$. □

Equations (7) and (8) ensure that there are no self-loops in the two networks. This is so that no baseless and circular arguments can occur. The acceptance of a sentence x cannot in itself be a reason for accepting x. Nor can the non-acceptance of x be a reason to accept x.

This result is not surprising from the perspective of a reason-based model. The interpretation is as follows. Take two agents with the same set of reasons who have seen the same evidence, and who have each constructed a story. If these stories are not identical, they must differ in the attitude they assign to at least two statements. If this were not the case, then the set of undisputed attitudes would be supporting both possible attitudes toward the sentence in dispute. One of the agents must not be justified in her attitude toward this sentence.

5 References

Baltzer, A. and N. Pennington (1983), "Reasoning About Conjunctions and Disjunctions of Events: An Explanation-Based Account, mimeo., Psychology Department, University of Colorado, Boulder.

Dekel, E., B. Lipman and A. Rustichini (1998), "Recent Developments in Modeling Unforeseen Contingencies", *European Economic Review*, 42.

Festinger, L. (1957), *A Theory of Cognitive Dissonance*, Row Peterson, Evanston, Illinois.

Hopfield, J. (1982), "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proceedings of the National Academy of Sciences*, April, 79.

Lam, R. (2001), "Learning Through Stories", mimeograph, Department of Economics, Northwestern University.

Pennington, N. and R. Hastie (1986), "Evidence Evaluation in Complex Decision Making", *Journal of Personality and Social Psychology*, 51(2).

Pennington, N. and R. Hastie (1988), "Explanation-Based Decision Making: Effects of Memory Structure on Judgement", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(3).

Pennington, N. and R. Hastie (1990), "Practical Implications of Psychological Research on Juror and Jury Decision Making", *Personality and Social Psychology Bulletin,* 16(1).

Pennington, N. and R. Hastie (1992), "Explaining the Evidence: Tests of the Story Model for Juror Decision Making", *Journal of Personality and Social Psychology,* 62(2).

Pennington, N. and R. Hastie (1993), "Reasoning in Explanation-Based Decision Making", *Cognition,* 49.

Rubinstein, A. (1998), *Modeling Bounded Rationality,* MIT Press, Cambridge, Massachusetts.

Savage, L. (1972), *The Foundations of Statistics,* 2nd Ed., Dover Press, New York, New York.

Shafir, E., I. Simonson, A. Tversky (1993), "Reason-Based Choice", *Cognition,* 49.