Local Thinking

Nicola Gennaioli and Andrei Shleifer[1]

First Draft, October 20, 2008

1. Introduction

Since the early 1970s, Daniel Kahneman and Amos Tversky (hereafter KT 1972, 1974, 1983, 2002) published a series of remarkable experiments documenting significant deviations from Bayesian theory of judgment under uncertainty. While KT's heuristics and biases program has survived substantial experimental scrutiny, models of heuristics have proved elusive[2]. In this paper, we offer a new model of decision making that accounts for quite a bit of this experimental evidence.

Perhaps the central heuristic unifying many of KT's ideas is representativeness, "defined as a subjective judgement of the extent to which the event in question is similar in essential properties to its parent population or reflects the salient features of the process by which it is generated" (Bar-Hillel 1982, quoting KT 1972, p 431). Judging probability by representativeness can trip a decision maker, as KT (1974) demonstrate. In one experiment, KT describe Jim as "shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure and a passion for detail." They then ask one group of subjects to estimate the probability that Jim is a farmer, sales-person, airline pilot, librarian, or physician, and another group of subjects which one of these Jim is most similar to. KT

[1] CREI and Harvard University, respectively. We are deeply grateful to Josh Schwarzstein for considerable input, and to Xavier Gabaix, Scott Kominers, David Laibson, Sendhil Mullainathan, Giacomo Ponzetto, and Jeremy Stein for extremely helpful comments.

[2] Partial exceptions include Mullainathan (2000) and Tversky and Koehler (1994), to which we come back.

find that the rankings of occupations by their probability is the same as that by their similarity, with librarian being the top choice. Moreover, and quite strikingly, the probability estimate of Jim being a librarian is basically the same whether subjects are told that Jim is drawn from a population of 70% librarians or that of 30% librarians. Subjects use similarity or representativeness to predict Jim's occupation.

In this paper, we present a nearly Bayesian model of decision making closely related to representativeness (indeed, Bayesian decision making is a limit case). In our model, an individual evaluates the likelihood of a hypothesis based on some partial evidence. In making this evaluation, the individual fills in from memory the missing details, which we call "frames." The question is how the frames are filled in. The idea of individuals filling in frames is consistent with KT's insistence that judgment under uncertainty is similar to perception. Just as an individual fills in details from memory when interpreting sensory data (for example, when looking at the duck-rabbit or when judging distance from the height of the object), in our model the decision maker recalls missing details for the hypothesis he is evaluating.

We make two assumptions about this process. First, in the spirit of KT, we assume that frames come to mind in order of their diagnosticity, which formally means their ability to predict the hypothesis being evaluated *relative* to other hypotheses. To take an example, suppose that a political candidate commits a blunder in a speech, and we wish to evaluate the probability that he is competent. We might quickly decide that this candidate is inarticulate, and therefore incompetent, even when most of the competent candidates are also inarticulate and make blunders. The lack of expressive ability is *diagnostic* of incompetence.

KT (2002, p.23) have a discussion of diagnosticity related to our model's definition: "Representativeness tends to covary with frequency: common instances and frequent events are generally more representative than unusual instances and rare events," but they add that "an attribute is representative of a class if it is very diagnostic; that is the relative frequency of this attribute is much higher in that class than in a relevant reference class." Most of our central results rely precisely on this distinction between diagnosticity and the relevant frequency (or likelihood) of frames.

Second, we assume that memory is limited, and not all potentially relevant frames come to mind. This assumption is essential because, with complete recall, the order of recall of missing data does not matter, and decision making is entirely Bayesian. We develop the implications of these two assumptions for judgement under uncertainty.

Our central results depend on the difference between the diagnosticity and likelihood of a frame. When, as is usually the case, the most diagnostic frames are also the most likely ones (most incompetent candidates are inarticulate, and most competent candidates are articulate), then local thinkers make judgment errors, but these errors tend to be modest. When, on the other hand, there is a mismatch between diagnosticity and likelihood of frames, and in particular the highly diagnostic frames for a given hypothesis happen to be highly unlikely, a local thinker's probability assessment becomes very poor. In particular, the probability of hypotheses whose most diagnostic frames are very unlikely can be heavily underestimated. Suppose that most candidates who make blunders, whether or not they are competent, are inarticulate, but that being articulate (and thus seldom making blunders) is more diagnostic of a competent candidate. The decision maker in this case severely underestimates the likelihood that a candidate is competent after a blunder. The mismatch between diagnosticities and likelihoods can also lead to substantial biases documented by KT.

One of the crucial puzzles that our model can account for is the conjunction fallacy. In perhaps their most famous experiment, KT described a young woman, Linda, as an activist in college, and asked their subjects about the relative likelihood of Linda's various activities today. Subjects reported that Linda is more likely today to be a bank teller *and* a feminist than just a bank teller, even though some bank tellers are surely not feminists. We show that the conjunction fallacy can be explained if details of "Linda" are filled in differently by a local thinker depending on what data he is given. In particular, if a former activist bank teller is represented with a diagnostic but very unlikely frame of a non-feminist, the local thinker can think that there are fewer such people around than there are formerly activist but now feminist bank tellers.

The next section of the paper presents our model, and section 3 characterizes its implications for evaluating uncertain hypotheses. Section 4 shows how the model can account for several phenomena documented by KT, including base rate neglect, failure of the conjunction rule, underestimation of implicit disjunctions, and insensitivity to predictability. Section 5 discusses possible extensions of our model.

2 The Model

An agent evaluates the probability of $N > 1$ exhaustive hypotheses $h_1,...,h_N$. The evaluation may occur in light of some data d. We call $\Pr^L(h_r|d)$ the local thinker's estimate of the probability of hypothesis h_r, r =1,..,N, which may differ from the true probability $\Pr(h_r|d)$. Our key premise is that the agent forms $\Pr^L(h_r|d)$ not on the problem's objective state space but on its mental representation produced by his background knowledge.

2.1 The Mental Space

The agent's background knowledge consists of a database *X* and a "recall" function $\pi : X \to [0,1]$. *X* collects all information in the agent's memory, represented by bundles $x = (x_1,..., x_K)$ of $K > 1$ dimensions along which the brain recognizes stimuli. Each dimension is discrete and finite, and we call X_i the domain of dimension *i* so that $X \equiv \prod_{i=1,...,K} X_i$. We assume that *X* is three dimensional, i.e. $K = 3$, but our results generalize to more dimensions. In this setting, each assessed hypothesis h_r and the data d correspond to a set of bundles (an event) in *X*. For each $r = 1,.., N$, $h_r \cap d \subseteq X$. That is, the set $h_r \cap d$ of bundles jointly satisfying the data and one hypothesis is a subset of *X*. Naturally, any hypothesis or data can be represented in *X*. When no data is given to the agent, $d = X$, so nothing is ruled out.

The function $\pi : X \to [0,1]$ measures the ease of recall of a bundle $x \in X$, where $\pi(x)$ is between 0 and 1. As suggested by memory research, the bundles stored in the brain are not all equally accessible to the agent. Some bundles may be easier to recall because the agent has experienced them more intensely in the past or because they are objectively more frequent. We focus on the case where $\pi(x)$ coincides with the true probability distribution of x, i.e. $\pi(x) = \Pr(x)$. This case shows the working of our model when memory operates with the correct prior. Before seeing how *X* and $\pi(x)$

produce mental representations, we illustrate our model with an example of an electoral campaign. We use this example throughout to illustrate our main results.

Example 1.A (Electoral Campaign): The Mental Space

An agent assesses the competence of a political candidate in light of a blunder. The first dimension of the mental space $X_1 \equiv \{\text{competent, incompetent}\}$ captures the candidate's competence, the second dimension $X_2 \equiv \{\text{articulate, inarticulate}\}$ captures his expressive ability. For simplicity we focus on a two dimensional version of this problem but we get back to three dimensions in example 1.G. The agent's background knowledge is represented by the table:

Blunder	articulate	inarticulate
competent	π_1	π_2
incompetent	π_3	π_4

Table 1.A

The entries in the cells are the probabilities of different bundles. Table 1.A lists the distribution of competence and expressive ability conditional on a blunder, so π_1 is the probability that a candidate who blundered is competent and articulate.

When the agent assesses whether the candidate is competent or incompetent, he assesses the two (alternative) hypotheses $h_1 \equiv \{x_1 = \textit{incompetent}\}$ and $h_2 \equiv \{x_1 = \textit{competent}\}$. As a consequence, h_1 (respectively h_2) identify in Table 1.A all bundles where an incompetent (respectively competent) candidate made a blunder, i.e. (*competent, articulate*), (*competent, inarticulate*) [respectively (*incompetent, articulate*), (*incompetent, inarticulate*)]. A hypothesis may fix two dimensions such as $h_1 \equiv \{x_1 = \textit{incompetent}, x_3 = \textit{articulate}\}$. In this case, h_1 perfectly identifies a bundle. For brevity, we identify sets with the dimensions they fix, calling for instance h_1 = *incompetent* (or *incomp.*) the hypothesis $h_1 \equiv \{x_1 = \textit{incompetent}\}$. We typically focus on the following numerical values:

Blunder	articulate	inarticulate
competent	0.2	0.2
incompetent	0.1	0.5

Table 1.A'

In Table 1.A', many blunders are made by inarticulate and incompetent candidates but quite a few competent and articulate candidates make blunders. Crucially, inarticulate candidates tend to be incompetent but many of them are actually competent. After the blunder, a Bayesian estimates Pr(*incomp*) = $\pi_3 + \pi_4 = 0.6$ and Pr(*comp*) = $\pi_1 + \pi_2 = 0.4$.◘

2.2 The Mental Representation and Probabilistic Assessments

The agent's represented state space is shaped by the recall of bundles in X prompted by the assessed hypothesis h_r, r = 1,…,N. Recall relies on two assumptions. First, a bundle is recalled earlier if it is more "diagnostic" of each given hypothesis h_r in a sense specified next. Second, limits in the agent's working memory or attention "truncate" the recall of bundles. These assumptions jointly imply that the represented state space is a selected subset of all bundles in X .

If $h_r \cap d = \{x'\}$ for some $x' \in X$, the agent perfectly represents h_r because, together with *d*, it identifies a unique bundle in *X*. For example, if h_r= *competent*, and d = (*articulate*), then the agent perfectly represents $h_r \cap d$ with (*competent, articulate*). The interesting issues arise when $h_r \cap d$ does not perfectly identify a bundle. For instance, the hypothesis that a blunder was made by an incompetent candidate leaves one missing dimension: the cause of blunder. The blunder may be the result of being inarticulate or the occasional mistake made by an articulate candidate.

More generally, for a given hypothesis h_r any bundle $x \in h_r \cap d$ can be used to represent it. In this setting we call "a frame" for hypothesis h_r a set $f \subseteq X$ such that $h_r \cap d \cap f = \{x'\}$ for some bundle $x' \in X$ and such that $\overline{h}_r \cap d \cap f \neq \phi$. Here $\overline{h}_r$ is the complement $\{x \in X\colon x \notin h_r\}$ of h_r. That is, a frame fully identifies a representation of hypothesis h_r, i.e. a $x \in h_r \cap d$, but is at the same time consistent with the alternative hypotheses (here summarized by $\overline{h}_r$). For instance, in the electoral campaign example, the hypothesis h_2= *competent* can be framed with the set f = {x_2 = *articulate*}, which intersects with h_1= *incompetent* at the bundle (*incompetent*, *articulate*) or with f = {x_2 =

inarticulate}, which intersects with h_1 at the bundle (*incompetent*, *inarticulate*). The term frame captures the idea that often the set *f* constitutes one possible specification of the details missing from the hypothesis and the data. For example, in the electoral campaign example, *f* = *articulate* is one way of specifying expressive ability, the characteristic left unspecified by hypothesis h_2= *competent*. In this respect, the term "frame" stresses the role of f in shaping the representation of a hypothesis in terms of a specific bundle in *X*.

A frame as defined above does not always exist. In particular, it does not exist if $h_r \cap d \cap f = \{x'\}$ can only be attained by choosing $f \subseteq h_r$ since in this case the intersection of *f* with $\overline{h}_r$ is always empty. This happens when hypothesis h_r does not specify an exact value for any of the dimensions in *X*. Assumption A.1 specifies what happens in this case. When instead h_r specifies the exact value of at least one dimension in *X* (such as in h_1= *incompetent*), then at least one frame exists for the hypothesis. In those cases, a frame constitutes one possible specification of the details missing from the hypothesis and the data. As in the electoral campaign example, there generally are several frames consistent with each $h_r \cap d$, and different frames bear different implications for the agent's evaluation. To aid the assessment of $\Pr(h_r|d)$, the brain starts to recall the various bundles in *X* consistent with $h_r \cap d$ by fitting f. Crucially, we assume:

A.1 (Recall Order): Fix data *d* and hypothesis h_r. Then, whenever it exists, frame f is recalled earlier in conjunction with hypothesis h_r the higher is its diagnosticity, namely its value according to the probability

$$\Pr(h_r|d \cap f) = \frac{\Pr(h_r \cap d \cap f)}{\Pr(h_r \cap d \cap f) + \Pr(\overline{h}_r \cap d \cap f)}. \tag{1}$$

When a frame for h_r does not exist, bundles $x \in h_r \cap d$ are recalled in random order.

Ties between frames can be broken randomly. A frame f is recalled earlier in conjunction with hypothesis h_r if – given d – this frame is more "diagnostic" of h_r, in the sense of being relatively more associated with it than with its altrenative $\overline{h}_r$.[3] If hypothesis h_r can be framed in M_r ways, expression (1) creates a ranking $k = 1,\ldots,M_r$ among them, with lower indexed frames f_r^k being easier to recall in conjunction with hypothesis h_r. In this ranking:

$$\begin{aligned} & f_r^1 = \arg\max_f \Pr(h_r|d \cap f) \\ & s.t. \quad h_r \cap d \cap f = \{x'\} \quad for \quad some \quad x' \in X \\ & \quad \overline{h}_r \cap d \cap f \neq \phi \end{aligned} \tag{2}$$

f_r^1 is the frame most diagnostic of h_r, the one immediately recalled by the agent.[4] The subscript r indicates that different hypotheses are generally represented using different frames. Expression (1) also implies that the recall of a frame is likely to depend on data d, but for notational simplicity we keep this dependence implicit.

Before interpreting expression (1), consider its application below.

[3] Little changes if one assumes that a frame is recalled earlier if it is more diagnostic of h_r relative to the other assessed hypotheses h_s, $s \neq r$. For the most part, we focus on the case where there are only two hypotheses, h_1 and $h_2 = \overline{h}_1$, in which case this distinction is immaterial.

[4] In light of (2), one can formally define f_r^k for every $k > 1$ as $f_r^k = \arg\max_f \Pr(h_r|d \cap f)$ such that $h_r \cap d \cap f = \{x'\}$ for some $x' \in X$, *and* $f \neq f_r^s$ for $s = 1,..,k-1$.

Example 1.B (Electoral Campaign): The Order of Recall

Suppose that the agent assesses $h_1 = incomp, h_2 = comp$ in light of a blunder. There are two ways to frame competence depending on expressive ability (articulate, inarticulate). To apply (1), notice that $h_2 = \bar{h}_1$. We thus obtain:

f	articulate	inarticulate
$\Pr(incomp \mid f)$	$\dfrac{\pi_3}{\pi_3 + \pi_1}$	$\dfrac{\pi_4}{\pi_2 + \pi_4}$
$\Pr(comp \mid f)$	$\dfrac{\pi_1}{\pi_3 + \pi_1}$	$\dfrac{\pi_2}{\pi_2 + \pi_4}$

Table 2: Diagnosticities of Frames

Thus, an incompetent candidate is immediately framed as inarticulate [formally, for $h_1 = incomp$ we have $f_1^1 =$(inarticulate), $f_1^2 =$(articulate)], when $\pi_4/\pi_2 > \pi_3/\pi_1$]. Intuitively, in this case incompetent candidates making blunders are relatively more frequent among inarticulate people, implying that the agent immediately associates incompetence with a lack of expressive ability. The same condition implies that the recall order for $h_2 = comp$ is the opposite, because when (as in the current case) there are two alternative (and exhaustive) hypotheses, a frame is more diagnostic for one of them if and only if is less diagnostic for the other. This is a key property of diagnosticity as it induces the agent to represent different hypotheses using different frames. In the numerical example of Table 1.A', the condition $\pi_4/\pi_2 > \pi_3/\pi_1$ holds since $\pi_4/\pi_2 = 5/2 > \pi_3/\pi_1 = 1/2$. Thus, in this numerical example after a blunder the agent represents an incompetent candidate as someone inarticulate, and the competent candidate as someone articulate. It is harder for the agent to recall that also articulate incompetents can make blunders and, most important, that inarticulate competent candidates can make blunders as well.◘

A crucial feature of A.1 is that different hypotheses tend to yield different orders of recalls of frames. In this important respect, our model is different from Mullainathan (2000), who considers how data are broadly framed in terms of categories. In our model, unlike in Mullainathan (2000), the hypothesis itself influences the frames used to evaluate it, at least in their order of recall. Diagnosticity of a frame (or in Mullainathan's framework, of a category) changes from one hypothesis to the next.

To belabour assumption A.1, notice that the most diagnostic frame f_r^1 must be sufficiently associated with hypothesis h_r, but need not maximize $\Pr(f|h_r \cap d)$, the likelihood of the frame given the data and the assessed hypothesis, which we later show to be a key factor affecting the accuracy of the agent's assessments. The order of recall implied by (1) thus captures the idea that the brain frames a hypothesis by selecting the missing data f most consistent with that hypothesis. In fact, many of our most interesting results, including our explanations of the biases resulting from heuristics, turn on the magnitude of the difference between the most likely frame and the most diagnostic one. Example 1.C below illustrates the difference between diagnosticity and likelihood.

Example 1.C (Electoral Campaign): Diagnosticity and Likelihood

The *likelihood* of frame *f* for *h_r* is measured by the probability $\Pr(f|h_r)$ that f occurs when h_r is true, as $\Pr(f|h_r)$ captures the share of the total probability of h_r accounted for by *f*. When the agent assesses $h_1 = incomp, h_2 = comp$ in light of a blunder, the likelihoods of the different frames for those hypotheses are:

f	articulate	inarticulate
$\Pr(f \vert incomp)$	$\frac{\pi_3}{\pi_3 + \pi_4}$	$\frac{\pi_4}{\pi_3 + \pi_4}$
$\Pr(f \vert comp)$	$\frac{\pi_1}{\pi_2 + \pi_1}$	$\frac{\pi_2}{\pi_2 + \pi_1}$

Table 3: Likelihoods of Frames

This implies that "inarticulate" is the most likely frame for incompetent if and only if $\pi_4 \geq \pi_3$ while "articulate" is the most likely frame for competent if and only if $\pi_1 \geq \pi_2$. This is indeed the case with the numbers of Table 1A' since $\pi_4 = 0.5 \geq \pi_3 = 0.1$ and $\pi_1 = 0.2 \geq \pi_2 = 0.2$. Thus, in our numerical example the most diagnostic and most likely frames coincide for both hypotheses. As we shall see, this aspect is important for the accuracy of the agent's assessment.

It is easy to see how diagnosticity and likelihood can become negatively related. Suppose for instance that we have the following distribution:

Blunder	articulate	inarticulate
Competent	0.02	0.38
Incompetent	0	0.6

Table 4

In this case, the overall probability of competence and incompetence are unchanged with respect to Table 1.A' and it is still the case that "articulate" is the most diagnostic frame for competence (as only competent candidates can be articulate). The main change is that now the bulk of candidates making blunders are inarticulate, irrespective of their competence. As a result, while most diagnostic, the "articulate" frame is the least likely for "competent".◘

In our setup, an agent representing a hypothesis recalls frames for that hypothesis in a certain order. This first step does not yet yield a model of mental representations. In the electoral campaign example, it seems important not only that after observing a blunder the agent thinks that the typical competent candidate is articulate and seldom makes blunders while the typical incompetent candidate is inarticulate and often makes blunders, but also that the event of a competent but inarticulate candidate escapes the agent's mind. Indeed, the order of recall does not matter for representations if all frames are recalled anyway. It is here that our second assumption of "local thinking" bites:

A.2. (Local Thinking): For each r =1,…,N and given d, the agent represents h_r by recalling only the first $b \geq 1$ bundles in the order f_r^k induced by expression (1).

Due to limits in the agent's working memory or attention, not all frames consistent with h_r are recalled by the agent to represent it. A frame is less likely to be integrated into the representation when it is harder to recall. Since $b \geq 1$, at least the bundle $h_r \cap d \cap f_r^1$ is retrieved from background knowledge. We call such a bundle "initial representation" of h_r. Two polar cases are of special interest: the case where $b = 1$, when thinking is fully local and only one frame is retrieved, and the case where

$b \geq M_r$ for each *r*, when the agent's representation is complete. In the campaign example, in considering whether a candidate is incompetent, a local thinker recalls only the inarticulate representation when the candidate commits a blunder, but both possible expressive abilities when there is greater memory capacity and $b = 2$.

A.2 starkly captures limited recall. If we alternatively assumed that the agent discounts the probability of hard to recall frames or that he recalls them with a certain probability, none of our substantive results would change. Mullainathan (2002) develops a model of limited memory where an agent recalls past events with a certain probability, but does not allow for different hypotheses to stimulate the recall of different scenarios or frames. Wilson (2002) is a normative study of decision making under bounded recall.

Consider how local thinking affects the estimation of probabilities. To assess the probability of h_s, the agent represents all hypotheses h_r, thus obtaining a mental representation of the problem's state space. In this space, the agent's assessment is:

$$\Pr^L(h_s | d) = \frac{\sum_{k=1}^{\min(b,M_s)} \Pr(h_s \cap d \cap f_s^k)}{\sum_{r=1,..,N} \sum_{k=1}^{\min(b,M_r)} \Pr(h_r \cap d \cap f_r^k)}, \tag{3}$$

where the assessed probability is the probability of the agent's *representation* of h_s relative to that of the agent's *representation* of all hypotheses h_r. That is, these probabilities are computed by considering for each hypothesis h_r only the first *b* frames f_r^k recalled in conjunction with it.

One important property of expression (3) is that the local thinker's assessment of the probability of a hypothesis depends on the other hypotheses examined in conjunction with it. This property is a direct consequence of imperfect recall and captures the idea that the description of the set of alternative scenarios can both remind the agent of scenarios in which a hypothesis is violated and also affect the framing of the alternative hypothesis by shaping the salience of alternative scenarios. In Section 4.3, we show that our model naturally yields a central assumption of Tversky and Koehler's (1994) support theory: the underestimation of the probability of implicit disjunctions.

We can rewrite (3) as:

$$\Pr^L(h_s|d) = \frac{\left[\sum_{k=1}^{\min(b,M_s)} \Pr(f_s^k|h_s \cap d)\right]\Pr(h_s \cap d)}{\sum_{r=1,..,N}\left[\sum_{k=1}^{\min(b,M_{rs})} \Pr(f_r^k|h_r \cap d)\right]\Pr(h_r \cap d)}, \tag{3'}$$

Expression (3') highlights the role of local thinking. If $b \geq M_r$ for all *r*, the agent perfectly represents the state space and his probability assessment is correct. The reason is that, since $\sum_{k=1}^{M_{rs}} \Pr(f_r^k|h_r \cap d) = 1$ must hold for all hypotheses h_r (i.e. the frames available in the mental space fully describe the hypothesis), expression (3') becomes equal to $\Pr(h_s \cap d)/\Pr(d)$, the Bayesian's (or correct) assessment of $\Pr(h_s|d)$. Biases in judgements can only arise when the agent's representation ability is limited.

For many of the examples we develop, we focus on the above formula when thinking is fully local, i.e. when *b* = 1. In this case,

$$\Pr^L(h_s|d) = \frac{\Pr(h_s \cap d \cap f_s^1)}{\sum_{r=1,..,N} \Pr(h_r \cap d \cap f_r^1)}. \tag{4}$$

Each h_r is represented by only taking into account the corresponding easiest to recall frame f_r^1. We now illustrate the above formulas in the electoral campaign example.

Example 1.D (Electoral Campaign): The State Space and Probability Assessments

An agent with b = 1 and the mental space of Table 1.A' assesses $h_1 = incomp, h_2 = comp$. Then, since Example 1.B showed that f_1^1 = *inarticulate*, f_2^1 = *articulate*, expression (4) yields:

$$\Pr^L(incomp) = \pi_4 /(\pi_4 + \pi_1) = 0.5/(0.5+0.2) = 0.71$$

After a blunder, the agent over-estimates the probability of the candidate being incompetent, whose correct value is $\Pr(incomp) = 0.6$. In this case, local thinking yields an over estimation of the probability of incompetence.

If instead the agent's mental space is the one of Table 1.A' we have that:

$$\Pr^L(incomp) = \pi_4 /(\pi_4 + \pi_1) = 0.6/(0.6+0.02) = 0.97$$

Now over-estimation of incompetence is huge. Example 1.F below discusses why.◘

The next section explores the circumstances under which local thinking yields biased or correct probabilistic assessments.

3. Biases in the Estimation of Probabilities

To examine how biases arise in our model, consider the simple setting where an agent assesses the odds of hypothesis h_1 relative to its alternative $h_2 = \overline{h}_1$. Both hypotheses concern the value of x_1, the first dimension of X which can take two values, i.e. $x_1 = \{h_1, h_2\}$. Data concern the second dimension x_2. The agent then recalls for each hypothesis r = 1,2 frames f_r^k in the order given by (1). In this case the total number of frames for h_1 and h_2 is the same, i.e. $M_1 = M_2 = M$ (this is always true if h_1 takes half of the elements of X_1). Expression (3) implies that the agent's estimate satisfies:

$$\frac{\Pr^L(h_1|d)}{\Pr^L(h_2|d)} = \left[\frac{\sum_{k=1}^{\min(b,M)} \Pr(f_1^k|h_1 \cap d)}{\sum_{k=1}^{\min(b,M)} \Pr(f_2^k|h_2 \cap d)}\right]\frac{\Pr(h_1|d)}{\Pr(h_2|d)} \tag{4}$$

The bracketed term is the key determinant of the agent's estimate, capturing the total likelihood the frames f_1^k recalled to represent h_1 relative to those f_2^k recalled to represent h_2. The agent correctly estimates the odds of h_1 relative to h_2 if and only if the bracketed term is equal to one. When this term is greater (respectively smaller) than one, the agent over (under) estimates the probability of h_1 relative to h_2.

When b =1, expression (4) becomes:

$$\frac{\Pr^L(h_1|d)}{\Pr^L(h_2|d)} = \left[\frac{\Pr(f_1^1|h_1 \cap d)}{\Pr(f_2^1|h_2 \cap d)}\right]\frac{\Pr(h_1|d)}{\Pr(h_2|d)}. \tag{5}$$

When thinking is fully local, the agent's assessment is correct if and only if the diagnostic frames used to represent h_1 and h_2 are equally likely for the respective hypotheses. When $\Pr(f_1^1|h_1 \cap d) > \Pr(f_2^1|h_2 \cap d)$, the agent represents h_1 with a frame that is relatively more likely than the frame with which he represents h_2. The agent thus over estimates the odds of h_1 because – for a given true odds ratio – his representation of h_1 is more likely than that of h_2. The opposite is true when $\Pr(f_1^1|h_1 \cap d) < \Pr(f_2^1|h_2 \cap d)$.

To evaluate the consequences of (4), call $F(t|h_r \cap d) = \sum_{k=1}^{t} \Pr(f_r^k|h_r \cap d)$ the cumulative likelihood of frames with recall order $k \leq t$ for hypothesis h_r. Define $E(k|h_r \cap d) = \sum_{k=1}^{M_r} k \Pr(f_r^k|h_r \cap d)$ as the average recall order for hypothesis h_r. $E(k|h_r \cap d)$ measures the extent to which easy to recall frames are also likely [in which case $E(k|h_r \cap d)$ is low], or unlikely [in which case $E(k|h_r \cap d)$ is high]. We can then show:

Proposition 1 If $b \geq M$ the agent's assessment is identical to that of a Bayesian. If $b < M$, the agent over (respectively under) estimates the odds of h_1 relative to h_2 if and only if $F(b|h_1 \cap d) > F(b|h_2 \cap d)$ (respectively $F(b|h_1 \cap d) < F(b|h_2 \cap d)$). The odds of h_1 relative to h_2 are (weakly) over-estimated for every $b < M$ if and only if $E(k|h_2 \cap d) \geq E(k|h_1 \cap d)$.

As we have seen before, if the agent thinks globally [i.e. $b \geq M$], he perfectly represents and thus assesses the hypotheses. This is the benchmark of full rationality. But even if thinking is local, the assessment is unbiased provided the represented state spaces capture the same share of their respective hypotheses' total probabilities [i.e. provided $F(b|h_1 \cap d) = F(b|h_2 \cap d)$]. Biases arise when the agent represents different hypotheses with frames of different likelihoods. We call this consequence of local thinking the "narrow focus" effect.

The narrow focus effect arises because, with local thinking, the recall intensity of different hypotheses affects the assessment of their probabilities; hypotheses that tend to be represented with unlikely frames are more likely to be underestimated. In particular, a hypothesis is always underestimated if the average recall order of its frames is higher than that of its alternative hypothesis. Intuitively, the representation of a hypothesis with lower average recall always includes relatively more likely frames than those of its alternative hypothesis. Before seeing under what conditions is the narrow focus effect is the strongest, we illustrate Proposition 1 with an example.

Example 1.E (Electoral Campaign): the Order of Recall and the Narrow Focus Effect

It is immediate to check that an agent with the state space of Table 1.A’ assessing $h_1 = incomp, h_2 = comp$ has:

	$t = 1$	$t = 2$
$F(t \mid incomp)$	0.83	1
$F(t \mid comp)$	0.5	1

Proposition 1 says that for *b* = *t* the odds of h_1 = *incomp* are overestimated if $F(t \mid incomp)$ > $F(t \mid comp)$, underestimated if $F(t \mid incomp)$ < $F(t \mid comp)$, and correctly estimated otherwise. Thus, the agent correctly assesses the odds of incompetence for *b* = 1. Since there are only two frames, for b = 2 the agent becomes Bayesian. Although in this example higher *b* trivially improves assessments, if there are more than two frames assessment bias can increase with *b* as additional frames may skew the represented space in favour of a hypothesis.◘

3.1 Diagnosticity, Likelihood, and The Narrow Focus Effect

Depending on the distribution $\Pr(x)$, local thinking may or may not generate assessment biases. This raises two questions: a) under what conditions on $\Pr(x)$ are those biases more severe?, and b) what is the role of the recall order assumed in A.1?

To answer these questions, notice that A.1 implies that the recall orders of two hypotheses h_1 and $h_2 = \overline{h}_1$ are negatively related. Formally, $f_1^k = f_2^{M+1-k}$ for $k = 1,...,M$. Indeed, with two alternative hypotheses, the most diagnostic frame for h_1 is the least diagnostic one for h_2 and vice-versa. This shows very clearly one key consequence of A.1: it induces the agent to represent different hypotheses with different frames. The intuition here is important: if the agent frames hypotheses by picking the best exemplar of each, he ends up representing these hypotheses with very different frames because best exemplars are naturally different. Proposition 1 illustrated under what conditions one hypothesis is systematically overestimated. We now dig into the workings of our model by highlighting a sufficient condition for that to happen:

Proposition 2. Denote by *k* the recall order of frames for h_1. Then, if $\Pr(f_1^k | h_1 \cap d)$ and $\Pr(f_1^k | h_2 \cap d)$ strictly decrease (respectively increase) in *k*, the agent over (respectively under) estimates the odds of h_1 relative to h_2 for every $b < M$.

The proof is in the Appendix. If frames that are more easily recalled under h_1 (i.e., those with low *k*) are also more likely for both hypotheses, then the under-estimation of the odds of h_2 is especially severe because h_2 is represented with its most diagnostic but least likely frames (those with high *k*) while h_1 is represented with its most diagnostic *and* likely frames (those with low *k*). The opposite bias occurs when, under both hypotheses, more likely frames are characterized by higher *k*.

The general principle here is that the narrow focus effect is particularly strong when diagnosticity and likelihood of frames are *positively* related for one hypothesis and *negatively* related for the other hypothesis. In this case, the first hypothesis is represented with a high probability frame, while the second with a low probability frame, giving rise to a strong narrow focus that leads to an over-estimation of the former hypothesis.

Although Proposition 2 does not imply that assessment biases only occur if diagnosticity and likelihood are far apart for one hypothesis, it suggests that in the latter case those biases are especially strong, as the result below confirms:

Corollary 1 Suppose $b = 1$. Then, consider the following two cases:

a) In the set of distributions $\Pr(x)$ such that $\Pr(f_1^k | h_1 \cap d)$ and $\Pr(f_1^k | h_2 \cap d)$ decrease in *k*, the maximal factor of over estimation of the odds of h_1 is arbitrarily large.

b) In the set of distributions $\Pr(x)$ such that $\Pr(f_1^k | h_1 \cap d)$ and $\Pr(f_1^{M+1-k} | h_2 \cap d)$ (weakly) decrease in *k*, the maximal factor of under (respectively over) estimation of the odds of h_1 is bounded above by *M*.

In a), the diagnosticity and likelihood of frames are negatively related for one hypothesis and positively related for the other hypothesis, so the narrow focus is very strong. In this case, if the most diagnostic frame for h_2 is very unlikely, the over estimation of h_1 can be huge. In b), where the diagnosticity and likelihood of frames are positively related for both hypotheses, biases are of limited (but possibly still large) magnitude. In this case, the largest over estimation of h_1 occurs when its likelihood is fully concentrated on its most diagnostic frame while hypothesis h_2 is fully spread among all of its *M* frames.

The above discussion highlights that greater dispersion of one hypothesis among disparate frames relative to its alternative causes the under estimation of the former hypothesis. Such over-estimation, however, is especially strong when the most likely frames are the same for both hypotheses. In this case, following the recall order of A.1, the agent stresses the hypotheses' differences rather than similarities, implying that only one hypothesis is represented with a likely frame, causing its over estimation. Biases are the strongest when the local thinker, looking for best exemplars, represents one hypothesis with an unlikely but distinctive frame, forgetting the commonalities between the hypotheses. We illustrate this idea in the example below.

Example 1.F (Electoral Campaign): Diagnosticity, Likelihood, and Narrow Focus

Table 1.A (and thus the estimate of Example 1.D) falls in case b) of Corollary 1, where for each hypothesis the most diagnostic and the most likely frames coincide and the bias is limited. The most likely frame for h_1 = *incomp* is "inarticulate", the most likely frame for h_2 = *comp* is "articulate", which are also the most diagnostic ones. The bias arises because conditional on a blunder competent candidates are more dispersed in expressive abilities than incompetent candidates, who are more concentrated on the inarticulate frame. Suppose instead we have:

blunder	articulate	inarticulate
competent	0.4	0
incompetent	0	0.6

Once more, the total probability of a competent and an incompetent candidate is the same as in Table 1.A', but only competent candidates are articulate and – most important – only incompetent candidates are inarticulate. In this case, diagnosticity and likelihood perfectly coincide. As a consequence, the local thinker does not lose any information by representing a competent candidate as articulate and an incompetent one as inarticulate. With no information loss, his assessment is exactly correct: the relative likelihood of h_1 and h_2 is 3/2.

In accordance with case b) of Corollary 1, when for both hypotheses the diagnosticity and likelihood of frames are positively related, the most severe bias arises when the probability of one hypothesis is fully concentrated on the most diagnostic frame while the probability of the other hypothesis is fully dispersed among all the frames. In the current example, this arises when π_4 = 0.6 and $\pi_1 = \pi_2$ =0.2, in which case the agent assesses the odds of h_1 relative to h_2 to be 3 and thus the over-estimation factor is equal to M = 2.

As illustrated in Example 1.D by the huge bias created by Table 4, over-estimation of incompetence can be much more severe when diagnosticity and likelihood are negatively related for one hypothesis but positively related for the alternative hypothesis, as stressed by case a) in Corollary 1. In the Example, this occurs when $\pi_1/\pi_3 > \pi_{2,}/\pi_4$ but $\pi_1/<\pi_2$ and $\pi_3< \pi_4$. Now, most candidates are inarticulate but the few articulate candidates are relatively more associated with competence. For example, suppose:

Blunder	articulate	inarticulate
Competent	ε	0.4-ε
Incompetent	0	0.6

where ε>0 is small. The articulate frame is diagnostic for h_1 = *competent* because only competent candidates are articulate. Yet, by representing h_1 = *competent* with an articulate candidate, the agent disregards that the bulk of competent candidates made a blunder *precisely because* they are inarticulate, which causes an over-estimation of incompetence. As $\varepsilon \to 0$, the over estimation bias of incompetence relative to competence, at 0.6/ε, becomes infinite, again consistent with Corollary 1.◘

We studied how biases can arise in our model when an agent assesses some hypotheses in light of some or no data. We have however abstracted from the question of how data provision can itself affect the agent's assessments. Although it is beyond the scope of this paper to systematically study the effects of data provision, this theme is sufficiently important to deserve some discussion. We return to it in Section 4.4.

4. Local Thinking and the Heuristics and Biases Program

We now illustrate with numerical examples how our model can account for several decision making biases related to representativeness uncovered by KT (e.g., 1974, 1983). We focus specifically on base rate neglect, failure of the conjunction rule, underestimation of implicit disjunctions, and insensitivity to predictability. We use the Linda experiment described in the introduction to illustrate the first two biases.

4.1 Neglect of Base Rates

Several experiments reveal subjects' failure to properly use base rates in the assessment of probability. For instance, KT (1974) gave subjects personality descriptions randomly sampled from a population of 100 professionals: engineers and lawyers. Subjects were told the total proportion of engineers and lawyers in the population but, in assessing the odds that a certain person was an engineer or a lawyer, they mainly focused on the personality description, barely taking the base rates of the two occupations into account. According to KT, subjects implicitly substituted the required assessment of probability with the more intuitive assessment of representativeness, judging the degree to which a personality description resembled the stereotype of an engineer or a lawyer.

Our model of local thinking can rationalize such neglect of base rates without requiring agents to employ non-probabilistic logic, but relying instead on the limited ability of subjects to represent or recall the underlying state space.

To restate the lawyer/engineer example in the Linda framework, suppose that Linda can be in one of two occupations, bank teller (BT) or social worker (SW), have two possible backgrounds, leftist activist (A) or non-activist (NA), and two possible current political orientations, feminist (F) or non-feminist (NF). The probability distribution of full descriptions of Linda is displayed in the table below, where τ and σ are the base probabilities of a bank teller and a social worker in the population, respectively.

A NA	F	NF
BT	$(2/3)(2\tau/8)$ $(1/3)(6\tau/8)$	$(1/3)(2\tau/8)$ $(2/3)(6\tau/8)$
SW	$(9/10)(2\sigma/3)$ $(1/2)(\sigma/3)$	$(1/10)(2\sigma/3)$ $(1/2)(\sigma/3)$

Table 5.

The numbers in Table 5 represent the probabilities of alternative scenarios. The numbers above the diagonal represent the joint probability distribution of Linda´s current political orientation and occupation conditional on having been an activist (A). The numbers below the diagonal represent this distribution for a non activist (NA).

Table 5 then captures two ideas. First, the bulk of bank tellers are former non-activists and currently non-feminist, and very few former activists subsequently become non-feminist bank tellers. Indeed, 6/8th of the base probability of bank tellers is captured by former non-activists, 2/3rd of whom are non-feminist. Of the remaining 2/8th of former activist bank tellers, only 1/3rd are non-feminist. Second, irrespective of whether they were activist or not, bank tellers tend to be relatively less feminist than social workers. In particular, 9 out of 10 formerly activist social workers are feminist, while only 2 out of 3 formerly activist bank tellers are feminist.

Suppose a subject is told that Linda was an activist (i.e., $d = \text{A}$) and asked to assess the probability she is a bank teller (BT). Suppose that the subject is a fully local thinker, i.e. $b = 1$. Then, a former activist bank teller is framed as a non-feminist (NF), while a former activist social worker is framed as a feminist (F). Concretely, for the hypothesis of bank teller, Table 5 implies that the diagnostic value $\Pr(BT|A, NF)$ of the "non-feminist" frame (NF) is larger than the diagnostic value $\Pr(BT|A, F)$ of the "feminist" frame (F). The latter diagnostic value is equal to $5\tau/(5\tau + 18\sigma)$, the former is equal to $5\tau/(5\tau + 4\sigma)$. The same calculation also implies that the most diagnostic frame for the hypothesis of a social worker is "feminist" F. Intuitively, in Table 5 former activists who became feminists are relatively more prevalent among social workers than among bank tellers.

These preliminaries imply that to a local thinker the odds of Linda being a bank teller are:

$$\frac{\Pr^L(BT|A)}{\Pr^L(SW|A)} = \frac{\Pr(BT, A, NF)}{\Pr(SW, A, F)} = \frac{(1/3)(\tau/4)}{(9/10)(2\sigma/3)} = \frac{5}{36}\frac{\tau}{\sigma} \tag{6}$$

The correct odds ratio is instead equal to:

$$\frac{\Pr(BT|A)}{\Pr(SW|A)} = \frac{3}{8}\frac{\tau}{\sigma} \tag{7}$$

If we compare (6) and (7), we see that the local thinker under-estimates the odds of Linda being a bank teller since he is less responsive than a Bayesian to the prior probability of those odds, τ/σ. Even if the odds of being a bank teller in the population τ/σ are quite high, after hearing the description of Linda the local thinker's updated probability under-weights the base rate by a factor of (3/8)*(36/5) = 108/40 relative to the correct Bayesian assessment, yielding neglect of the base rate τ/σ.

The intuition for this result is simple. The evidence *d* = A skews the agent's mental representation in favour of "social worker," since the fact of Linda being a former activist allows the local thinker to represent her occupation as "social worker" with the most likely frame (activist, feminist), while it prompts him to represent the bank teller with a low likelihood frame (activist, non-feminist). As in the Example 1.G, this results from the local thinker's tendency to represent a hypothesis (bank teller) with the most diagnostic frame (non-feminist) rather then with the most likely one (feminist).

This reasoning also suggests that dropping the description of Linda as a former activist should improve the assessed odds of her being a bank teller because, while this change does not affect the framing of a social worker (still depicted as a former activist and now feminist), it would lead the local thinker to represent the bank teller with the more likely frame of a former *non-activist* (and now non-feminist), yielding

$$\frac{\Pr^L(BT)}{\Pr^L(SW)}=\frac{\Pr(BT,NA,NF)}{\Pr(SW,A,F)}=\frac{(2/3)(6\tau/8)}{(9/10)(2\sigma/3)}=\frac{5}{6}\frac{\tau}{\sigma}, \tag{8}$$

an almost correct unconditional probability assessment. Of course, it is not always the case that data provision reduces the quality of judgment. In the current example, this happens because the data provision has an asymmetric effect on the two hypotheses. Since being a former activist is fully consistent with being a social worker but mildly inconsistent with being a bank teller, the provision of *d* = *A* reduces the local thinker's ability to represent the latter hypothesis with a likely frame, causing its underestimation.

4.2 Violations of the Conjunction Rule

The Linda example is most famous for the violation of the conjunction rule, which states that the probability of a conjoined event C&D cannot exceed the probability of each event C or D by itself. We illustrate how our model accounts for violation of the conjunction rule when data that Linda is a former activist is provided, in Section 4.2.1. Section 4.2.2 accounts for conjunction rule violations with no data given.

4.2.1 Conjunction Rule Violation with Data

Our model of local thinking can shed light on conjunction rule violations with parameter values from Table 5. As we saw previously, when *d* = A is given to the subject, he represents a bank teller with the frame "non-feminist" and a social worker with the frame "feminist". The new hypothesis of "feminist bank teller" leaves no gaps to be filled and is thus correctly represented with the frame "former activist, now feminist bank teller", i.e. (BT, A, F). As a result, the local thinker estimates:

$$\frac{\Pr^L(BT|A)}{\Pr^L(BT,F|A)}=\frac{\Pr(BT,A,NF)}{\Pr(BT,A,F)}=\frac{(1/3)(2\tau/8)}{(2/3)(2\tau/8)}=\frac{1}{2}<1 \tag{9}$$

The local thinker grossly violates the conjunction rule.

Interestingly, in this example, the conjunction rule would not be violated if the data *d* =A was not given to the subject. In this case, the event "bank teller" would be framed with "formerly non-activist, now non-feminist" [i.e. (NA, NF)], yielding:

$$\frac{\Pr^L(BT)}{\Pr^L(BT,F)}=\frac{\Pr(BT,NA,NF)}{\Pr(BT,A,F)}=\frac{(2/3)(6\tau/8)}{(2/3)(2\tau/8)}=3>1 \tag{10}$$

Again, the intuition is straightforward once this result is viewed in light of Example 1.G. The data *d* = A given to the local thinker induce him to frame the hypothesis of "bank teller" with a now moderate but former activist person. The problem, however, is that the probability of this outcome is very low because, conditional on being a former activist, most bank tellers are feminists.

In our model, the violation of the conjunction rule comes from the separation between frames' diagnosticity and likelihood highlighted in Proposition 2 and Corollary 1. The conjunction rule can only be violated in situations where the subject represents the broader hypothesis with a very low likelihood frame. If the subject represents a hypothesis with the most likely frame, then the conjunction rule can never be violated. In our example, the local thinker represents the hypothesis of Linda being a bank teller with the more numerous category of formerly activist, now feminist bank tellers rather than with the unlikely category of formerly activist now non-feminist bank tellers.

In this respect, our model deals with the question, "why don't people realize that the population of bank tellers includes the feminist ones?", by replacing it with the question more pertinent to describing a local thinker, "why don't people represent a former activist, now bank teller with the more *likely* frame of "feminist?". The answer is in Example 1.G: the local thinker never considers that a bank teller can be a feminist because feminist is a characteristic disproportionately associated with social workers, and does not therefore match with the image of an exemplar bank teller.

Crucially, in the absence of data, the state space is no longer skewed against the hypothesis of bank teller because the local thinker now selects the frame of a former non activist. As a consequence, since in Table 5 formerly non activist and now non-feminist bank tellers are more likely than formerly activist and now feminist bank tellers, in (10) the conjunction rule is not violated.

Conjunction rule violations hinge on the separation between diagnosticity and likelihood of frames, not on data provision per se; indeed, such violations can also arise when data are not given, as Section 4.2.2 shows. Before moving on, we wish to stress that one interpretation of the conjunction error holds that instead of assessing $\Pr(h|d)$, subjects intuitively assess the inverse probability $\Pr(d|h)$ of the data given the hypothesis in question. Thus, in the Linda experiment subjects confuse the probability $\Pr(BT|A)$ of Linda being a bank teller with the probability $\Pr(A|BT)$ that a bank teller is Linda, and the probability of Linda being a feminist bank teller with the probability $\Pr(A|F,BT)$ that a feminist bank teller is Linda.[5] This error can yield the conjunction fallacy because being feminist can increase the chance of being Linda. Indeed, in our example of Table 5:

$$\Pr(A|BT) = 1/8 < \Pr(A|F,BT) = ¼.$$

One shortcoming of this explanation of the conjunction fallacy is that it does not elucidate the thought process by which the subject substitutes the target assessment $\Pr(h|d)$ with the delivered assessment $\Pr(d|h)$. Interestingly, our model can shed some light on why the agent may assess $\Pr(d|h)$ instead of $\Pr(h|d)$. This is best seen by writing:

$$\frac{\Pr(h_1|d)}{\Pr(h_2|d)} = \frac{\Pr(d|h_1)}{\Pr(d|h_2)} \frac{\Pr(h_1)}{\Pr(h_2)} \tag{11}$$

In the above expression, the subject may mistakenly estimate the odds of h_1 given the data with the odds of the data given h_1 if he does not account for the base rates of the two

[5] In a personal communication, Xavier Gabaix proposed a “local prime” model complementary to our local thinking model. Such model exploits the above intuition about the conjunction fallacy. Specifically, in the local prime model an agent assessing h_1, …, h_n evaluates $\Pr^{L'}(h_i|\mathrm{d}) = \Pr(\mathrm{d}|h_i)/[\ \Pr(h_1|\mathrm{d}) + \ldots + \Pr(h_n|\mathrm{d})]$.

hypotheses. Insofar as our model can account for the subject's failure to properly account for base rates, it can also induce the subject to estimate $\Pr(h_1|d)/\Pr(h_2|d)$ as if he were estimating $\Pr(d|h_1)/\Pr(d|h_2)$. This can be directly seen in our numerical example where, according to expression (9), after observing d = A the local thinker assesses the odds of the bank teller versus feminist bank teller to be 1/2. But this is exactly equal to the value of Pr(A|BT)/Pr(A|F,BT), which is 1/2 as well.

More broadly, KT (1983) themselves discussed the possibility that subjects may confuse Pr(h|d), with Pr(d|h), and demonstrated the inapplicability of this explanation to a range of situations where the conjunction fallacy is detected. For example, after being told that the tennis player Bjorn Borg reached the Wimbledon final, subjects were asked to assess whether it was more likely that in the final Borg would lose the first set or whether he would lose the first set but win the match. Most subjects violated the conjunction rule by stating that the second outcome was more likely than the first. Although our model is consistent with this experimental result, the hypothesis that subjects assess Pr(d|h) is not. In the Borg example, the two hypotheses are "losing the first set" and "losing the first set but winning the final," and the data is "Borg reached the final." Both hypotheses perfectly predict the data, so Pr(d|h) = 1 for both hypotheses. This is clearly not what the subjects estimate.

There is another family of conjunction rule violations that is inconsistent with subjects confusing Pr(h|d) with Pr(d|h), and that is when the conjunction rule is violated in the absence of data provision. The next section studies this case in detail.

4.2.2 Conjunction Rule Violations in the Absence of Data

We propose a numerical example tracking an experiment carried out to document conjunction rule violations without data [e.g. see Kahneman and Tversky (1983)]. Suppose an agent is asked to estimate the probability of the event “A massive flood somewhere in North America in which more than 1000 people drown” and of the event “An earthquake in California causing a flood in which more than 1000 people drown”. The correct probability of the second event should clearly be not larger than that of the first because the latter event is included in the former.

Suppose that the agent’s mental space has three dimensions: the type of the flood, which can either be severe (S) or mild (M), the cause of a flood, which can either be a earthquake (E) or a tornado (T), and the location of the flood, which can either be California (C) or the rest of North America (NC). Suppose that the distribution in X has the following features:

M S	E	T
C	$(1-x)e_C$ xe_C	$t_C/2$ $t_C/2$
NC	$e_{NC}/2$ $e_{NC}/2$	$(1-z)t_{NC}$ zt_{NC}

Table 6

In Table 6, e_C and t_C are the probabilities of an earthquake and a tornado in California, respectively, e_{NC} and t_{NC} are the probabilities of an earthquake and a tornado in the rest of north America, respectively, and $x > 1/2$ and $z > 1/2$ are respectively the share of earthquakes causing severe floods in California and of tornados causing severe floods in the rest of North America. Obviously, parameter values should be such that $e_C + t_C + e_{NC} + t_{NC} = 1$.

Table 6 has two key features. First, in the rest of North America earthquakes are sufficiently milder than in California that they cause fewer severe floods. Indeed, while only 1/2 of earthquakes cause severe floods in North America, in California x > 1/2 earthquakes cause severe floods. Second, tornados are sufficiently milder in California than in the rest of North America that they cause fewer severe floods. Indeed, while only 1/2 of tornados cause severe floods in California, z > 1/2 tornados cause severe floods in the rest of North America. The specific assumption that *x*, z > 1/2 is not important, what matters is that earthquakes and tornadoes are stronger in California and in the rest of North America, respectively. We make the natural assumption $z > x$, meaning that tornados are more likely to cause severe floods than earthquakes.

These assumptions imply that a local thinker represents a mild flood (M) with the frame (C,T). Indeed, Table 6 implies $\Pr(M|T,C) = \Pr(M|E,NC) = 1/2$ > $\Pr(M|E,C) = 1-x$ > $\Pr(M|T,NC) = 1-z$. Most important, since the least diagnostic frame for a mild flood (M) is the most diagnostic one for the severe flood (S), the above ranking implies that a local thinker represents the event of a severe flood (S) with the frame (T, NC), i.e. as a severe flood caused by a tornado in the rest of North America

The event "Severe flood caused by an earthquake in California" need not be framed because it uniquely identifies the bundle (S, C, E). Given this framing, the local thinker assesses the odds of (S, C, E) relative to (S) as:

$$\frac{\Pr^L(S)}{\Pr^L(S,C,E)} = \frac{\Pr(S,NC,T)}{\Pr(S,C,E)} = \frac{zt_{NC}}{xe_C} \qquad (11)$$

As long as the prior probability of disastrous earthquakes in California is sufficiently high relative to that of disastrous tornados in the rest of North America, (i.e. provided $zt_{NC} > xe_C$), the subject violates the conjunction rule in this case as well, even though he is given no data at all. This bias arises because the subject represents severe floods with the diagnostic but unlikely frame (S, NC, T), forgetting that severe floods caused by earthquakes in California also belong to the set of severe floods.

4.3 Underestimation of Implicit Disjunctions

Fischoff, Slovic and Lichtenstein (1979) asked car mechanics, as well as lay people, to estimate the probabilities of different causes of a car's failure to start. They document that on average the probability assigned to the residual hypothesis – "The cause of failure is something other than the battery, fuel system, or the engine" – went up from 0.22 to 0.44 when that hypothesis was broken up into more specific causes (e.g. the starting system, the ignition system). Respondents – including most remarkably experienced car mechanics – discounted hypotheses that were not explicitly mentioned. The under-estimation of implicit disjunctions such as residual hypotheses relative to explicit disjunctions has been documented in many other experiments and is the cornerstone of Tversky and Koehler's (1994) support theory.

By incompletely representing hypotheses, a local thinker is naturally predisposed to underestimate implicit disjunctions. Furthermore, conjunction rule violations – which Section 4.2 has shown to be a possible consequence of local thinking – themselves reveal how the probability of an implicit disjunction such as bank teller is underestimated relative to that of a constituent event such as feminist bank teller. As we have seen, conjunction rule violations only arise in our model under certain conditions, particularly when the diagnosticity and likelihood of frames are far apart. It is thus interesting to ask under what conditions underestimation of implicit disjunctions arises in our model.

Suppose that a local thinker with $b = 1$ assesses the probability of two hypotheses h_1 and h_2 without observing any data (it is immediate to generalize the analysis to the case where some data is observed). The local thinker then assesses

$$\Pr^L(h_1) = \frac{\Pr(h_1 \cap f_1^1)}{\Pr(h_1 \cap f_1^1) + \Pr(h_2 \cap f_2^1)} \tag{12}$$

Suppose instead that the same local thinker assesses three hypotheses h_1, $h_{2,1}$ and $h_{2,2}$ where $h_{2,1} \neq \phi$, $h_{2,2} \neq \phi$, $h_{2,1} \cup h_{2,2} = h_2$ and $h_{2,1} \cap h_{2,2} = \phi$. Thus, he must now separately assesses the two disjoint constituent elements $h_{2,1}$ and $h_{2,2}$ of h_2 rather than the implicit disjunction h_2. In this second problem, the local thinker recalls a frame $f_{2,1}^1$ for hypothesis $h_{2,1}$ and a frame $f_{2,2}^1$ for hypothesis $h_{2,2}$, thereby estimating:

$$\Pr^L(h_1) = \frac{\Pr(h_1 \cap f_1^1)}{\Pr(h_1 \cap f_1^1) + \Pr(h_{2,1} \cap f_{2,1}^1) + \Pr(h_{2,2} \cap f_{2,2}^1)} \tag{13}$$

Notice that h_1 is still represented with f_1^1 because unpacking the implicit disjunction h_2 does not affect expression (1) for h_1.

It is immediate to see that the implicit disjunction is strictly under-estimated relative to the explicit disjunction when (12) is larger than (13), which in turn boils down to the condition:

$$\Pr(h_{2,1} \cap f_{2,1}^1) + \Pr(h_{2,2} \cap f_{2,2}^1) > \Pr(h_2 \cap f_2^1)\,. \tag{14}$$

The implicit disjunction is underestimated when the probability of the bundle with which such a hypothesis is represented is smaller than the sum of the probabilities of the bundles with which the explicit disjunction is represented. The only way in which (14) may not be fulfilled is if the explicit disjunction evokes two diagnostic frames $f_{2,1}^1$ and $f_{2,2}^1$, both of which are much less likely than f_2^1. A sufficient condition for this is:

$$h_2 \cap f_2^1 \in \left\{ h_{2,1} \cap f_{2,1}^1, h_{2,2} \cap f_{2,2}^1 \right\} \tag{15}$$

The implicit disjunction is under-estimated when the representation of its unpacked version always contains the representation of the implicit disjunction (plus one or more additional bundles, depending on the number of explicit disjunctions).

In our model, condition (15) is always satisfied, for two reasons. First, since $h_{2,1}$ and $h_{2,2}$ are disjoint [i.e. $h_{2,1} \cap h_{2,2} = \phi$], the agent represents the implicit disjunction h_2 by choosing a bundle belonging to either $h_{2,1}$ or $h_{2,2}$ but not to both. Second, and given the previous fact, one can rewrite the agent's recall process of a frame for h_2 as:

$$\begin{aligned} & f_2^1 = \arg\max_f \Pr(h_{2,1} \mid f) + \Pr(h_{2,2} \mid f) \\ & s.t. \quad h_{2,1} \cap f = \phi \quad or \quad h_{2,2} \cap f = \phi \quad but \quad not \quad both \end{aligned} \tag{16}$$

In sum, the agent represents the implicit disjunction h_2 by selecting the most diagnostic representation among all of its explicit disjunctions. Hence, the representation of the explicit disjunction always includes the representation of the implicit disjunction, implying that in our model condition (15) always holds.

Local thinking, then, naturally yields underestimation of implicit disjunctions. Such underestimation occurs because unpacking of a hypothesis h_2 into its constituent events preserves the diagnosticity of the former's representation while allowing the agent to integrate into the representation bundles that would not be recalled otherwise.

These implications of local thinking can account for the car mechanic experiment. Suppose the different causes of a car's failure to start can be represented by $X \equiv \{battery, fuel, ignition\}$, where *fuel* stands for "fuel system" and *ignition* stands for "ignition system" and $\Pr(x) > 0$ for every cause of the failure to start. The agent is initially asked to assess the likelihood that the car's failure to start is for a reason other than battery troubles. That is, he is asked to assess the likelihood of hypotheses $h_1 = \{fuel, ignition\}$, $h_2 = battery$. A local thinker arbitrarily represents the implicit disjunction $h_1 = \{fuel, ignition\}$ by one of its constituent elements *fuel* or *ignition* (e.g., he may represent reasons other than battery troubles by ignition system troubles or by fuel system troubles). Indeed, this is one case where a frame as defined in Section 2 does not exist. As a result, according to assumption A.1 the agent randomly chooses a bundle in $h_1 = \{fuel, ignition\}$. Without loss of generality, suppose that the local thinker represents h_1 by *fuel*. The local thinker then attaches probability

$$\Pr^L(h_1) = \frac{\Pr(fuel)}{\Pr(fuel) + \Pr(battery)} \tag{17}$$

to the cause of the car's failure to start being a reason other than *battery*.

Now suppose that the implicit disjunction h_1 is broken up into its constituent elements, $h_{1,1}$ = *fuel* and $h_{1,2}$ = *ignition* (e.g., the individual is asked to separately assess the likelihood that the car's failure to start is due to ignition troubles or to fuel system troubles). Clearly, the local thinker represents $h_{1,1}$ by *fuel* and $h_{1,2}$ by *ignition*. As before, the agent represents the other hypothesis h_2 by *battery*. As a consequence, the local thinker now attaches greater probability to the car's failure to start being for a reason other than the battery because:

$$\begin{aligned}&\Pr^L(ignition)+\Pr^L(fuel)=\frac{\Pr(ignition)+\Pr(fuel)}{\Pr(ignition)+\Pr(fuel)+\Pr(battery)}\\&>\Pr^L(h_1)=\frac{\Pr(fuel)}{\Pr(fuel)+\Pr(battery)}\end{aligned}\tag{18}$$

4.4 Effects of Data Provision

A cursory look at expression (4) shows that data provision can sway assessments by affecting the local thinker's representation of the hypotheses (the term in square brackets). This effect occurs even if the data are barely predictive or objectively irrelevant, namely when the conditional and unconditional odds of h_1 relative to h_2 coincide. Section 4.2.1 already illustrated how data provision may trigger the violation of the conjunction rule. We now look more systematically at the role of data by considering insensitivity to predictability.

Various studies show that people often fail to account for the reliability of the evidence used in making probabilistic judgements, which are often heavily shaped by scarcely informative data. Our model provides one way to think about this kind of bias. Consider now the role of data in our electoral campaign example.

Example 1.G (Electoral Campaign): The Role of Data.

Suppose that an agent assesses $h_1 = incomp$ versus $h_2 = comp$ without observing the candidate's speech (i.e. whether he committed a blunder or not). Suppose that the third dimension of the mental space is $X_3 \equiv \{\text{homerun, blunder}\}$ and captures the quality of the candidate's speech. Suppose furthermore that the three dimensional mental space is:

homerun blunder	articulate	inarticulate
Competent	0.25 0.1	0.05 0.1
Incompetent	0.1 0.05	0.1 0.25

Table 7

With these parameter values, the total probability of a blunder is one half, the distribution conditional on blunder is the same as that of Table 1.A' and articulate and competent candidates are associated together and make fewer blunders than inarticulate and incompetent candidates.

Now each hypothesis has four possible frames, each consisting of a combination of speech quality and expressive ability. Using the number of Table 7 we can easily find that

f	good team, blunder	bad team, blunder	good team, homerun	bad team, homerun
$\Pr(incomp \mid f)$	1/3	5/7	2/7	2/3
$\Pr(comp \mid f)$	2/3	2/7	5/7	1/3

In this case, without observing the candidate's speech, a local thinker with b = 1 represents a competent candidate as an articulate one delivering homeruns and an incompetent candidate as an inarticulate one making blunders. As a result, the local thinker assesses:

$$\Pr^{L}(incomp) = 0.25/(0.25+0.25) = 1/2,$$

which is identical to the correct prior probability of a candidate being incompetent. Data provision in the form of observing the candidate's blunder induces a bias in favour of the hypothesis of incompetence. It is easy to see that the bias turns to favour the hypothesis of competence if the agent observes a homerun. Below we discuss the reason for this result. ◘

The above Example shows how local thinkers over-react to data. After a blunder, the local thinker inflates the odds of incompetence, exaggerating the predictive power of blunders. Crucially, such over-reaction occurs because the blunder is inconsistent with the "articulate" frame used by the local thinker to represent competent candidates, and thus reduces the perceived likelihood of such an "articulate" frame. By representing a competent candidate's blunder as the occasional mistake of an articulate candidate, the local thinker disregards that some competent candidates are inarticulate and make blunders, causing him to grossly over-estimate the odds of incompetence.

The general lesson is that data provision can exacerbate biases by reducing the likelihood of the frame with which one of the two hypotheses is represented. In this case, which occurs when data are negatively correlated with the frame used to represent one of the two hypothesis, the narrow focus effect becomes stronger. Ironically, in this case scarcely informative data can greatly affect the agent's assessments by increasing the *informational loss* associated with local thinking. The conjunction rule violation of Section 4.2.1 was due to the same likelihood-reducing effect of data.

We now illustrate this reasoning with an example inspired by the experimental evidence on subjects' insensitivity to low predictability. In one study, subjects were presented with several paragraphs, each describing the performance of a student-teacher during a particular practice lesson. Some subjects were asked to evaluate the quality of the lesson, other subjects were asked to predict the standing of each student-teacher five years after the practice lesson. The judgements made under the two conditions were identical, irrespective of subjects' awareness of the limited predictability of teaching competence five years later on the basis of a single trial lesson.

To see how local thinking may be responsible for this bias, consider the similar example of a candidate giving a job talk in an academic department. A subject is asked to assess the quality of the candidate (or the probability that the candidate is tenured in the department) based on the quality of the talk. As usual, suppose there are three dimensions, the first is the quality of the candidate which can be high or low, the second is the quality of the talk, which can be good or bad, while the third is the training of candidate, which can be good or poor. Suppose that the mental space is as follows:

good talk	Poor train	Good train
high quality	0	1/4
low quality	ε	1/4 – ε

Table 8.A

bad talk	Poor train	Good train
high quality	1/4 - ε	ε
low quality	1/4	0

Table 8.B

Here ε > 0 is assumed to be small. In this example, the quality of the talk is completely uninformative about the candidate's quality [as Pr(high quality| good talk) = Pr(high quality| bad talk) = 1/2]. This is admittedly extreme, but it highlights the mechanism by which uninformative data can affect judgement. The two key features of the above table are the following. First, the candidate's training is highly correlated with the quality of the talk: the bulk of bad talks are given by poorly trained candidates, while the bulk of good talks by well trained ones. Second, the candidate's training is mildly informative of his or her quality as Pr(high quality| good training) = $\frac{1/4 + \varepsilon}{1/2}$> Pr(high quality| poor training) = $\frac{1/4 - \varepsilon}{1/2}$.

Consider now how a full local thinker (i.e. *b* =1) reacts to uninformative data such as the quality of the talk. Since poorly trained candidates tend to be of lower quality than well trained ones, poor training is diagnostic for low quality, inducing the local thinker to represent a high quality candidate as a well trained one and a low quality candidate as a poorly trained one. As a consequence, conditional on observing a good talk or a bad talk the local thinker assesses:

$$\frac{\Pr^L(highquality \mid goodtalk)}{\Pr^L(lowquality \mid goodtalk)} = \frac{\Pr(highquality, goodtalk, goodtrain)}{\Pr(lowquality, goodtalk, poortrain)} = \frac{1/4}{\varepsilon}$$

$$\frac{\Pr^L(highquality \mid badtalk)}{\Pr^L(lowquality \mid baddtalk)} = \frac{\Pr(highquality, badtalk, goodtrain)}{\Pr(lowquality, badtalk, poortrain)} = \frac{\varepsilon}{1/4}$$

The local thinker grossly over-estimates the quality of the candidate after a good talk and under-estimates the quality of the candidate after a bad talk, even when the quality of the talk is completely uninformative about the candidate's true quality.

After hearing a good talk, the local thinker argues that a low quality candidate is poorly trained and would have thus delivered a bad talk, and forgets that many well trained candidates giving good talks are of low quality. Upon hearing a bad talk, the local thinker believes that a high quality candidate is well trained and would have thus delivered a good talk, and forgets that a few poorly trained candidates giving bad talks are of high quality.

As discussed above, this judgement bias is due to the fact that, although per se uninformative, the data are inconsistent with the way one of the two hypotheses is framed, in the sense that the data reduce the likelihood of one of the hypotheses' frame. In the job talk example, the assessment bias is much stronger than in the electoral campaign example. The intuition is that now the data and the frame are highly correlated. The bulk of bad talks are given by poorly trained candidates and the bulk of good talks are given by well trained ones, irrespective of the candidates' quality. In this case, data sharply separate the diagnosticity and the likelihood of frames for one of the hypotheses, thereby greatly exacerbating assessment biases.

To summarize, we have seen that in our model the agent over-reacts to uninformative data when the data is consistent with the way one hypothesis is framed and inconsistent with the way its alternative is framed. The extent of such over-reaction is especially strong when the data and the frame are highly correlated.

5. Conclusion.

Incorporating some features of representativeness into a nearly Bayesian model of decision making can account for many of the biases documented by Kahneman and Tversky. Most importantly, both the conjunction and the disjunction fallacies emerge quite naturally from our model.

Our analysis raises the question of whether other features of representativeness, as well as the other key heuristics, namely availability and anchoring, can be incorporated into our model. Our model clearly does not cover them. Availability focuses on the ease of recall, which presumably would generally yield a different recall order than diagnosticity. With anchoring, recall is shaped by irrelevant factors priming the mind.

At the same time, it is worth pointing out that these additional heuristics, at a broad level, share some crucial similarities with representativeness as we model it. Specifically, these heuristics share with our model the feature that something other than the data primes the decision maker's model of the world, as well as the feature that memory is imperfect. In our model, it is the hypothesis itself that primes the frame, and a local thinker ignores some of the alternative frames. With anchoring, the priming is done by an irrelevant anchor. And of course limited memory is central to availability. We hope to pursue these similarities among heuristics in future work.

6. Proofs

Proof of Proposition 1. The proof immediately follows from the fact that the condition $F(b|h_1 \cap d) > F(b|h_2 \cap d)$ which ensures that the odds of h_1 relative to h_2 are over-estimated for every $b < M$ effectively says that $F(b|h_2 \cap d)$ first order stochastic dominates $F(b|h_1 \cap d)$, which is in turn equivalent to $E(k|h_2 \cap d) \geq E(k|h_1 \cap d)$.

Proof of Proposition 2. By definition, at any given $b < M$ the hypothesis h_1 is represented with frames $\left(f_1^k\right)_{k \leq b}$ while hypothesis h_2 is represented with frames $\left(f_1^{M+1-k}\right)_{k \leq b}$. As such, the odds of h_1 are over-estimated at *b* if and only if

$$\sum_{k=1}^{b} \Pr(f_1^k|h_1 \cap d) \geq \sum_{k=1}^{b} \Pr(f_1^{M+1-k}|h_2 \cap d) \tag{19}$$

Suppose now that $\Pr(f_1^k|h_1 \cap d)$ and $\Pr(f_1^k|h_2 \cap d)$ strictly decrease in *k*. Then, one can easily show that the above condition is met for every *b* < *M*. Suppose in fact that for a certain $b^* < M$ the above condition is not met. That is, suppose that

$$\sum_{k=1}^{b^*} \Pr(f_1^k|h_1 \cap d) < \sum_{k=1}^{b^*} \Pr(f_1^{M+1-k}|h_2 \cap d) \tag{20}$$

Then, at some $b^{**} \leq b^*$ it must be the case that $\Pr(f_1^{b^{**}}|h_1 \cap d) < \Pr(f_1^{M+1-b^{**}}|h_2 \cap d)$. But then, since $\Pr(f_1^k|h_1 \cap d)$ and $\Pr(f_1^k|h_2 \cap d)$ strictly decrease in *k*, it must also be the case that $\Pr(f_1^b|h_1 \cap d) < \Pr(f_1^{M+1-b}|h_2 \cap d)$ for all $b \leq b^*$. The same property implies that $\Pr(f_1^b|h_1 \cap d) < \Pr(f_1^{M+1-b}|h_2 \cap d)$ for all $b > b^*$. But then, this implies that (20) must hold for all $b > b^*$, including *b* = *M*, which is inconsistent with the fact that:

$$\sum_{k=1}^{M} \Pr(f_1^k|h_1 \cap d) = \sum_{k=1}^{M} \Pr(f_1^{M+1-k}|h_2 \cap d) = 1 \tag{21}$$

must necessarily hold. Hence, if $\Pr(f_1^k|h_1 \cap d)$ and $\Pr(f_1^k|h_2 \cap d)$ strictly decrease in *k* condition (19) must always hold and the odds of h_1 are always (weakly) overestimated. By using the same logic, it is immediate to show that the odds of h_2 are always (weakly) overestimated when $\Pr(f_1^k|h_1 \cap d)$ and $\Pr(f_1^k|h_2 \cap d)$ strictly increase in *k*.

Proof of Corollary 1. Take the set of distributions $\Pr(x)$ such that $\Pr(f_1^k|h_1 \cap d)$ and $\Pr(f_1^k|h_2 \cap d)$ decrease in *k*. Then, Proposition 1 implies that in this class of distributions the odds of h_1 are over-estimated at any *b* < *M*. In particular, the factor of over-estimation for a local thinker at *b* = 1 is equal to $\Pr(f_1^1|h_1 \cap d) / \Pr(f_1^M|h_2 \cap d)$. Then, consider a distribution whereby $\Pr(f_1^1 \cap h_1 \cap d) = \Pr(h_1 \cap d)\left[1 - \varepsilon^2 \dfrac{1-\varepsilon^{M-1}}{1-\varepsilon}\right]$ and

$\Pr(f_1^k \cap h_1 \cap d) = \Pr(h_1 \cap d)\varepsilon^{2(k-2)}$ for all k≥2 and $\Pr(f_1^1 \cap h_2 \cap d) = \Pr(h_2 \cap d)\left[1 - \frac{1-\varepsilon^{M-1}}{1-\varepsilon}\right]$ and $\Pr(f_1^k \cap h_2 \cap d) = \Pr(h_2 \cap d)\varepsilon^{(k-2)}$ for all k≥2, where 0 < ε < 1. Under this distribution, lower indexed frames are more diagnostic of h_1 because the probability of bundles belonging to h_1 decays much faster with *k* than that of bundles belonging to h_2. Under both hypotheses, however, the probability of bundles decreases in *k*, which implies that this probability distribution belongs to the class where $\Pr(f_1^k | h_1 \cap d)$ and $\Pr(f_1^k | h_2 \cap d)$ decrease in *k*. Notice that when b =1 the odds of h_1 relative to h_2 are equal to:

$$\frac{\Pr(f_1^1 \cap h_1 \cap d)}{\Pr(f_1^1 \cap h_2 \cap d)} = \frac{\Pr(h_1 \cap d)}{\Pr(h_2 \cap d)} \frac{\left[1 - \varepsilon^2 \frac{1-\varepsilon^{M-1}}{1-\varepsilon}\right]}{\varepsilon^{M-2}}$$

For given true odds ratio $\frac{\Pr(h_1 \cap d)}{\Pr(h_2 \cap d)}$, the estimated odds ratio becomes infinite as ε → 0.

Consider part b) of the corollary. If $\Pr(f_1^k | h_1 \cap d)$ and $\Pr(f_1^{M+1-k} | h_2 \cap d)$ (weakly) decrease in *k*, the two hypotheses are represented with their most likely frames. Thus, the greatest over estimation of h_1 relative to h_2 is reached when $\Pr(f_1^1 | h_1 \cap d) = 1$ and $\Pr(f_1^M | h_2 \cap d) = 1/M$. That is, when h_1 is concentrated on its representation while the distribution of h_2 is fully dispersed among all frames. In this case, the agent over estimates the odds of h_1 by a factor of *M*. Accordingly, when $\Pr(f_1^1 | h_1 \cap d) = 1/M$ and $\Pr(f_1^M | h_2 \cap d) = 1$, the agent under estimates the odds of h_1 by a factor of *M*. To conclude, notice that in those distributions it is indeed the case that *k* indicates the recall order for h_1 because in both cases the diagnosticity of a frame for h_1 falls in *k*.

References

Fischhoff, Baruch., Paul Slovic, and Sarah Lichtenstein. 1978. “Fault Trees: Sensitivity of Assessed Failure Probabilities to Problem Representation. *Journal of Experimental Psychology: Human Perceptions and Performance*, 4, 330-344.

Kahneman, Daniel, and Amos Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.

_____. 1974. “Judgment Under Uncertainty: Heuristics and Biases,” *Science*, 185, 1124-1131.

_____. 1983. “Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment,” *Psychological Review*, 91, 293-315.

Maya Bar-Hillel. 1982. “Studies of Representativeness,” in D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgement under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Mullainathan, Sendhil. 2000. “Thinking through Categories,” mimeo.

_____. 2002. “A Memory-Based Model of Bounded Rationality,” *Quarterly Journal of Economics*, 117(3), 735-774.

Tversky, Amos, and Derek Koehler. 1994. “Support Theory: A Nonextensional Representation of Subjective Probability,” *Psychological Review*, 101, 547-567.

Wilson, Andrea. 2002. “Bounded Memory and Biases in Information Processing,” mimeo.