2012

Information Projection: Model and Applications

Kristof Madarasz, *London School of Economics and Political Science*

Information Projection: Model and Applications

KRISTÓF MADARÁSZ
London School of Economics

First version received January 2010; Final version accepted September 2011 (Eds.)

Abstract

People exaggerate the extent to which their information is shared with others. This paper introduces the concept of such *information projection* and provides a simple but widely applicable model. The key application describes a novel agency conflict in a frictionless learning environment. When monitoring with ex-post information, biased evaluators exaggerate how much experts could have known ex-ante and underestimate experts on average. Experts, to defend their reputations, are too eager to base predictions on ex-ante information which substitutes for the information jurors independently learn ex-post, and too reluctant to base predictions on ex-ante information which complements the information jurors independently learn ex-post. Instruments which mitigate Bayesian agency conflicts are either ineffective or directly backfire. Applications to defensive medicine are discussed.

Keywords: Biased Beliefs, Asymmetric Information, Hindsight Bias, Monitoring, Defensive Medicine.

1 Introduction

Economic analysis commonly assumes that people understand the extent to which their information is private. Evidence from several important domains shows however that people systematically mispredict informational differences. In particular, people too often think as if others knew what they did, and too often act as if others could guess their private information correctly.

Investigations after accidents – such as the collapse of Space Shuttle Challenger in '86 – not only collect novel information about what went wrong, but cause people to exaggerate how easy it would have been to predict the tragedy ex-ante. Given the ex-post information, it is possible to understand how NASA officials should have interpreted ex-ante facts in a way to avoid the disaster. What many investigators then typically fail to fully appreciate is that such insights are only possible in hindsight.

The same mistake is prevalent in communication: speakers are cursed by their expertise and exaggerate how much of their information on a subject is shared with their audiences. As a result, software engineers design interfaces, manufacturers provide user-manuals and professors discuss material in ways that their intended audiences find too hard to follow.

To understand such behavior more carefully, this paper introduces a simple but widely applicable model of such *information projection*: people are aware of informational differences but project their private information, exaggerating the extent to which others have access to the content of their private information.

Section 3 of the paper presents the model. I consider a general environment where people receive signals about an underlying state privately. Having observed a signal s, a person exaggerates the probability with which others have also observed the content of this signal. A person who projects information in this fashion fails to appreciate the extent to which others need to think and act without the information she has. Through a clear identifying property, the model ties together a set of empirically well-documented social mispredictions – hindsight bias (Fischhoff, 1975), curse-of -knowledge (Camerer, Loewenstein and Weber, 1989) or the illusion of transparency (Gilovich, Medvec and Savitsky 1998) – and provides a framework to study their consequences in a unified manner.

The main application of the model is to the classic problem of performance evaluation. Examples are ubiquitous: corporate boards need to learn about the talent of CEOs, governments about the effectiveness of their intelligence agencies or their police force, patients about the quality of physicians etc. I show that information projection here gives rise to a novel form of agency conflict. Biased evaluators using ex-post information will underestimate the quality of experts. Experts in turn will have incentives to engage in specific defensive practices.

To illustrate this agency conflict, consider a medical example. Radiologists differ in skill; even when all have access to the same radiograph, the best ones hardly ever miss a tumor when it is visible on the X-ray while bad ones often do. Initially, all radiologist recommend a treatment based on an ambiguous X-ray s_0. Once the recommendation is adopted and results are in, an evaluator reviews the case.

The evaluator now has access to information s_1 which was unavailable ex-ante: interim medical outcomes are realized and new X-rays are ordered. A small tumor is typically difficult to spot on an initial X-ray. When the location of a major tumor is known ex-post, however, all radiologists have a much better chance of finding the small one on the original X-ray. A biased evaluator thinks as if such ex-post information had already been available ex-ante. By projecting this information, the evaluator is thus too surprised when observing that the treatment failed and too often interprets a success to be the norm. Hence she underestimates the radiologist's competence *on average*.

Evidence from law and medicine suggests that professionals do fear such underestimation and respond to it strategically. In medicine, the fear of the 'retrospectroscope' refers to exactly such anticipation. To describe the possible strategic responses, let's return to the medical example. In addition to the X-ray, the radiologist now has to decide what other tests s_0' to run. He knows that independent of his choice evaluators will learn some novel information s_1 ex-post. How does the anticipation that evaluators will project s_1 when assessing his competence ex-post alter the radiologist's incentives to produce an additional test ex-ante?

An ex-ante test substitutes for the evaluator's ex-post information if knowing its outcome

decreases how much additional information s_1 reveals ex-post. A second MRI on a broken bone contributes less to a successful diagnosis than the first one. If ordering an MRI ex-post is a likely event, the radiologist will have an incentive to order an MRI ex-ante even if the initial X-ray would have been medically sufficient. In this manner, physicians might take too many biopsies, expose patients to too much radiation, or recommend surgery too often so that their ex-ante predictions appear less incompetent in hindsight.

An ex-ante test instead complements the ex-post information if knowing s_0' increases how much additional information s_1 reveals ex-post. Suppose a social worker can make a valuable, but still ambiguous, phone call to a foster family. Evaluators will ultimately always learn whether a child was injured, but this can only confirm child abuse once combined with the content of the ex-ante phone call. Ironically, the social worker will enjoy a higher professional reputation if she does not make the phone call ex-ante. In this manner, engineers will be reluctant to record equivocal forecasts of catastrophes, and physicians will be reluctant to order vague radiological charts if they fear that ex-post biased evaluators will interpret these pieces of information as clean ex-ante evidence.

The above responses describe an agency conflict where increased ex-post scrutiny or stronger reputational incentives induce agents to distort the composition of the ex-ante produced information from complement to substitute signals. Solutions that correct for standard agency conflicts are either ineffective or may directly backfire. Instead, limiting monitoring can help restore efficiency. I specify conditions under which reputations formed on the basis of monitoring successful cases more intensely than failed cases mitigate a radiologist's incentives to engage in the above defensive practices.

The predictions of the model are consistent with evidence from *defensive medicine,* defined as medical practices adopted to minimize false liability rather than maximize cost-effective health care. Studdert et al. (2005) show, that to protect their reputation, a vast majority of physicians engage in both the over-production and under-production of skill-intensive medical tests. Kessler and McClellan (1996, 2000) show that weaker liability pressure lowers medical costs but not the over-all quality of care, and that this occurs primarily through changes in diagnostic practices.

The paper is structured as follows. Section 2 summarizes evidence for the model which is presented in Section 3. Sections 4 describes the agency conflict. Section 5 discusses the robustness of the results and how selective monitoring can help mitigate this conflict. Section 6 provides a second application of the model to communication. Here the model provides a rationale for the use of communication protocols to improve information transmission and fairness at the workplace. Section 7 concludes.

2 Evidence

In this Section, I review evidence from both the lab and the field. The source of the informational asymmetry is greatly varied across the studies and while individual studies are subject to multiple interpretations, the sum-total of the evidence provides a compelling case

for the model.[1]

Informational asymmetries are inevitable when people learn over time. Here projection bias implies that those exposed to superior information in the present exaggerate how well others in the past could have predicted their current information. In a between-subject design, Fischhoff (1975) showed that no matter which outcome of an uncertain event was reported ex-post, this significantly increased how likely people thought that this event would occur ex-ante. Such *hindsight bias* is the most widely documented mistake in the judgment literature. Hundreds of studies using various paradigms and debiasing techniques provide overwhelming support for Fischhoff's original findings. For recent surveys see Rachlinski (1998) or Guilbault et al. (2004).

A significant number of studies with professionals confirm the prevalence of hindsight bias outside the lab. Among many others, Arkes et al. (1981) document significant hindsight bias in a sample of 75 practicing physicians. In a study with 121 anesthesiologists judging actual medical cases, Caplan et al. (1991) document the bias in ruling ex-ante negligence. In law, Anderson et al. (1997) demonstrate strong effects with practicing judges, and Hastie et al. (1999) show dramatic hindsight bias with jury eligible citizens. Baron and Hersey (1988) demonstrate the same effect when the ex-post information is simply the outcome of an objective lottery and the ex-ante choice is between monetary gambles.

Simultaneous exchanges are also ripe with informational asymmetries and evidence confirms projection in these domains as well. Newton (1990) provides an illustrative study in the context of communication. Subjects were randomly assigned to be tappers or listeners. Tappers privately picked a song from a public list and tapped out its rhythm in front of the listeners. Out of the 120 songs tapped, listeners correctly identified only 2.5%. After tapping, tappers were asked to predict these odds and the mean prediction was around 50%. Information projection here explains a twenty-fold overestimation. Informed bystanders who knew the selected song, but only listened to the tapping, also vastly overestimated the success rate (Heath and Staudenmayer, 2000), while uninformed bystanders did not, (Keysar and Henly, 2002). The same effects were established in direct verbal communication, (Keysar and Henly, 2002), and email communication, (Kruger et al., 2005).

The quality of the tapper's action directly affects informational differences in these experiments, but evidence suggests that people project private information in exchanges where differences are fixed and people do not need to judge the quality of information transmission. In a study by Loewenstein, Moore and Weber (2006), subjects had to spot the difference between two pictures. They were divided into three groups: an uninformed group with subjects who received no further information, an informed group with subjects who were told the difference, and a third group with subjects who could choose to learn the difference for a fee. In all treatments, subjects had to guess the fraction of people who would correctly identify the difference in the *uninformed* condition. Subjects were paid for the accuracy of their predictions.

The true fraction was 20%. As Figure 1 indicates, informed subjects greatly overestimated this fraction relative to uninformed subjects. A significant proportion of subjects

[1] For a more extensive discussion of the evidence, see Madarász (2009).

Information condition		Mean prediction	Standard deviation	N
Uninformed		30.1 %	25.6	66
Informed		58.2 %	32.7	66
Choice		40.6 %	29.5	66
Choice (unopened)	(71%)	34.6 %	29.0	47
Choice (opened)	(29%)	55.4 %	25.8	19

Figure 1: Loewenstein, Moore, and Weber (2006).

paid to learn the difference, producing more biased estimates (55% versus 30%) and hence systematically lowered their earnings. Thus people not only acted as if they projected their private information, but also paid for information that only biased their judgments.[2]

In the context of financial markets, Camerer, Loewenstein and Weber (1989) provide support for information projection. Subjects traded assets via a double-oral auction in two distinct markets. In the first, subjects were only told the past performance of the traded companies and asset returns were determined by current performance. In the second, subjects learned actual earnings as well, but here asset returns were determined by the prices established in the first market. The prices in the second market were biased by 30% towards the actual earnings and individual judgments were biased by 60%. Traders with a smaller bias traded more aggressively, as if anticipating the bias of others. Camerer (1992) – also a precursor to LMW – confirmed the findings of CLW, but showed that some subjects learned to avoid biasing information. When traders had the option to bid for extra information, bids started high, but converged to zero after a few repetitions. A few subjects even bid negative, indicating some awareness of potential information projection that is hard to reverse or resist.

Finally, a set of other psychological mispredictions indicates that people project private information about their internal states as well. Gilovich, Medvec and Savitsky (1998, 2000) show that people greatly overestimate the probability that their emotions are detected by others or that their lies, once made, will be discovered.

2.1 Related Literature

To the best of my knowledge, the current paper is the first to introduce a model of information projection. The closest to my paper is the study of CLW (1989) who introduce an incomplete model of anchoring to explain their data. Although both their model of anchored expectations and the model of information projection explain these specific results, anchored expectations are generally inconsistent with information projection. A person who is anchored to his own beliefs would typically violate information projection and vice versa. I

[2]Similar results were established for non-visual tasks such as the solution of a logical puzzle. To control for curiosity, LMW told subjects that they would learn the solution to the logical and visual puzzles at the end of the experiment.

formally demonstrate this in the Appendix.

Biais and Weber (2009) complete the anchoring approach of CLW and offer a model of intrapersonal hindsight bias. There people correctly remember the variance of their past beliefs but misremember the mean. The authors specify conditions under which this leads to under-reacting to financial news and offer empirical support using investment bankers from London and Frankfurt. Mangelsdorff and Weber (1998) also follow the anchoring approach of CLW and provide a formal example of hindsight bias in the context where an agent chooses between binary lotteries.[3]

This paper also complements a recent literature on limited strategic reasoning in Bayesian games, (Eyster and Rabin, 2005), (Koessler and Jehiel, 2008). In these models, people's average beliefs are correct and people predict information differences correctly, but fail to appreciate how much others condition their actions on their true private information. More broadly, this paper also relates to cognitive hierarchy approaches to strategic thinking, (Camerer, Ho and Chong, 2004) or (Crawford, Costa-Gomes and Iriberri 2010), where people have simplified models of the depth of reasoning by others.

3 Model

Consider an environment where people privately observe signals about the payoff-relevant state $\omega \in \Omega$. A signal is a function from the finite set of states to the set of lotteries over a realization space, $s_j : \Omega \to \Delta Z$. Signals are interpreted given a common prior σ_0 over the state space. There are N signals and M people. Person k's actual information is the set of signals S^k whose realizations she knows. The set of all possible information sets is $\mathbb{N}$ which is the power set of N.

To characterize the distribution of information, let $p_j^k \in (0, 1]$ denote the initial probability that person k observes the realization of signal s_j. Let us collect these probabilities over signals and across people into a vector $\mathbf{p} = \{\{p_j^k\}_{j=1}^N\}_{k=1}^M$. This vector $\mathbf{p}$ describes the true distribution of information in the environment. Each sub-vector $\mathbf{p}^k = \{p_j^k\}_{j=1}^N$ is a probability distribution over $\mathbb{N}$. An unambiguous increase in person's k information is a change in the sub-vector p^k which weakly increases each of its components. In turn, the informational environment can be summarized by the tuple $\Gamma = \{\Omega, \sigma, \{s_j\}_{j=1}^N, p\}$.

3.1 Definition

The vector $\mathbf{p}$ determines the true distribution of signals and corresponds to rational expectations. Information projection introduces a bias in these expectations. A biased person exaggerates the probability that the signals whose realizations she knows are also in the information set of others. I introduce a parameter $\rho \in [0, 1]$ to express the degree of such mistaken information projection.

Definition 1 *Person k with information set S^k exhibits information projection of degree*

[3] Camerer and Malmendier (2007) also argue about the importance of attribution errors in organizations

$\rho > 0$ *if her perception of person* $i's$ *information distribution is given by* $p^i(\rho)$ *where*

$$p^i_j(\rho) = (1-\rho)p^i_j + \rho \text{ if } s_j \in S^k \text{ and } p^i_j(\rho) = p^i_j \quad \text{if } s_j \notin S^k \tag{1}$$

By projecting information, a person misperceives the distribution of information. Under full information projection, $\rho = 1$, a person believes that all the information she knows is shared with others. In the case of partial information projection, $0 < \rho < 1$, she believes that the probability that her information is available to others is between the truth and the full projection case. Finally, $\rho = 0$ corresponds to correct expectations.[4] Some remarks are in order:

- The degree of projection in above is uniform across signals. This is for notational simplicity only and the model immediately extends to heterogenous projection. There, the degree of projection is a vector, rather than a scalar. Formally, if ρ^k_j is the degree to which person k projects signal j, then collecting these terms into a vector ρ^k is person k's generalized degree of projection. Here an increase in the bias is an increase in *any* component of this vector. All results extend to heterogenous projection.

- The definition is formulated without explicit reference to time. If a different identity is assigned to the past or future selves of person i, then the definition claims that a biased person projects her current information onto the past or future selves of person i. Thus a biased person exaggerates the extent to which person i has observed the content of her information in the past or that he will know/remember her information in the future.

- The definition adopts a simple linear form, but it is sufficient to assume that $p^i_j(\rho)$ continuously increases in ρ in the interval $[p^i_j, 1]$.Any functional form which satisfies this will lead to the same predictions as those described in this paper.

- While the literal meaning of the definition is that people exaggerate the probability that others observe their signals, in a more general interpretation people exaggerate the probability that the information content of their signals is reflected in the information available to others. Here projecting information is not a false belief about how public information transmission is but rather the extent to which one's own views are the reflection of truly private information.

- It is both intuitive and psychologically realistic that some types of information are more likely to be projected than others. While the model provides no a prior predictions on the degree to which a particular piece of information is subject to projection relative to another piece, it provides a potentially useful framework to establish such claims

[4]The model can be extended by allowing p to depend on the state ω and hence to be state-dependent random variables rather than state-independent parameters. In this formulation, $p(\omega)$ is a projector's updated belief about the distribution of information after receiving signals. The definition then can be applied to this updated vector $p_i(\omega)$ in the same way as above. The model thus can be interpreted as one where people have heterogenous priors. Importantly though, relative to postulating the existence of heterogenous priors with no theory of the way these priors will be heterogenous, the current model makes clear *directional* predictions on people's conflicting estimates as a function of the true informational environment.

empirically. Importantly, for the purposes of this paper, the way in which information differences in Γ are partitioned into different signals will matter neither for the key properties of the model nor for the main results derived in the context of the applications. As I show this in Section 3.3, both of these will depend only on the fact that $\rho \geq 0$.

3.2 A Dinner Party

To illustrate the model, I turn to a simple example. Tanya is invited for dinner. Alex is her host and either prepares fish or meat. If Alex is kind, his goal is to prepare Tanya's favorite. If selfish or unkind, he cares only about his own taste. Tanya and Alex know their own tastes, but not the other party's. Initially, Alex is equally likely to be kind or selfish and independently, he is equally likely to prefer fish or meat.

Alex privately receives a noisy signal about Tanya's taste. This signal conveys her true preference over fish and meat $\frac{2}{3}$ of the time, and the false one $\frac{1}{3}$ of the time. The table below summarizes the inferences of a Bayesian and a fully biased Tanya.

	Bayesian, $\rho = 0$	Fully biased, $\rho = 1$
$\pi_1^\rho(\theta_{kind} \mid x = \text{ right dish}) =$	$\frac{2/3+2/3}{2/3+1+2/3} = \frac{4}{7}$	$\frac{1+1}{1+1+1} = \frac{2}{3}$
$\pi_1^\rho(\theta_{kind} \mid x = \text{wrong dish}) =$	$\frac{1/3+1/3}{1/3+1+1/3} = \frac{2}{5}$	$\frac{0}{1} = 0$
$E\pi_1^\rho(\theta_{kind}) =$	$\frac{7}{12}*\frac{4}{7}+\frac{5}{12}*\frac{2}{5} = \frac{1}{2}$	$\frac{7}{12}*\frac{2}{3} = \frac{7}{18}$

Tanya makes two types of inferential mistakes: *over-inference* and *underestimation.* First, she overinfers kindness when served the meal she likes, and hostility when served the meal she dislikes. Second, Tanya underestimates Alex's kindness on average.

To understand the first effect, note that Tanya wrongly believes that Alex knows her taste for sure. She thus too often concludes that good intentions result in good outcomes and bad intentions in bad outcomes, and as a result, misattributes differences in tastes to differences in social intentions. On average, she attributes too much kindness to Alex if the two have the same taste, and too much hostility to Alex if their tastes differ. Numerically,

	$\rho = 0$	$\rho = 1$
$E\pi_1^\rho(\theta_{kind} \mid \text{similar taste}) =$	$\frac{5}{6} * \frac{4}{7} + \frac{1}{6} * \frac{2}{5} = \frac{19}{35}$	$\frac{5}{6} * \frac{2}{3} = \frac{5}{9}$
$E\pi_1^\rho(\theta_{kind} \mid \text{different taste}) =$	$\frac{1}{3} * \frac{4}{7} + \frac{2}{3} * \frac{2}{5} = \frac{16}{35}$	$\frac{1}{3} * \frac{2}{3} = \frac{2}{9}$

To understand the second effect, note that the two overinference distortions don't cancel each other out. The excess negative attribution is always larger than the excess positive attribution. Since a biased Tanya exaggerates how often a kind Alex serves the right dish, she expects to be served the dish she likes too often and underestimates Alex's kindness on average. Such underestimation is a key mechanism for this paper. Similar calculations show that because Tanya also exaggerates how often Alex serves the right dish if their tastes coincide, she also underestimates the similarity their tastes an average.

In short, Tanya misattributes differences in tastes to differences in social intentions and underestimates both the kindness and the similarity of Alex on average. While the ex-ante expected Bayesian posterior is uniform, the biased one is $\{\theta_{ks} = \frac{7}{36}, \theta_{kd} = \frac{7}{36}, \theta_{us} = \frac{7}{36}, \theta_{ud} = \frac{15}{36}\}$. Tanya and Alex might then too often depart as foes rather than friends.

3.3 Information and Underestimation

Let's return to the definition. As mentioned before, a key identifying property of the model is determined alone by the claim that $\rho > 0$. Below I show that this is true because, independent of further details of the informational environment, by exaggerating the probability that Alex observes her signals, Tanya exaggerates how much Alex knows about the state. To make the above statement precise, suppose Alex maximizes a bounded and state-dependent von Neumann-Morgenstern utility function over a finite action-set given information $S \subset \mathbb{N}$. Formally, his expected payoff is:

$$u^*(S) = \max_y E_\sigma[u_A(y, \omega) \mid S] \tag{2}$$

Since Alex may receive signals privately, from Tanya's perspective $u^*(S)$ is a random variable distributed over $\mathbb{N}$. A rational Tanya has correct beliefs about this distribution. A fully biased Tanya instead wrongly believes that Alex maximizes his expected utility after combining the signals he truly observes with whatever signals she observes. Since more information always allows one to achieve a greater expected utility, Tanya misperceives this distribution and exaggerates how well Alex can achieve his objectives.

Let $\pi^\rho(u^*) \in \Delta\mathbb{N}$ be a privately informed ρ-biased Tanya's belief about the distribution of Alex's expected payoff. The next result shows that a biased Tanya always exaggerates Alex's expected payoff and this exaggeration is increasing in her information.

Lemma 1 *For all* $\mathbf{p}$, $\pi^\rho(u^*)$ *first-order stochastically dominates* $\pi^{\rho'}(u^*)$ *if* $\rho > \rho'$. *If* $\mathbf{p} = \widehat{\mathbf{p}}$ *except that* $\mathbf{p}^{Tanya} \geq \widehat{\mathbf{p}}^{Tanya}$, *then* $\pi^\rho(u^*)$ *fosd* $\widehat{\pi}^\rho(u^*)$ *for all* $\rho > 0$.

This lemma helps identify the model.[5] First, it shows that greater information projection leads to a greater exaggeration of a projectee's payoff. Second it shows that holding a projector's bias constant, increasing a projector's information leads to greater exaggeration. Since the comparative static is based purely on increasing the projector's information, the

[5] The proof of this lemma is an application of the classic result of Blackwell (1953) on the comparison of information sets. While Blackwell (1953) offers only a partial ordering of information sets, the above result shows that misperceptions induced by information projection can be ordered by his criterion.

above result depends neither on the details of how information is distributed nor on how informational differences are partitioned into signals. Specific restrictions on the environment Γ, motivated by economic or psychological considerations, might be imposed on the signal structure and such restrictions will enrich the predictions of the model, but will not change the above result.[6]

Lemma 1 is useful to establish a general consequence of information projection to inference problems. In the dinner example, Tanya underestimated both the kindness and the similarity of Alex. To establish a similar result more generally, I consider a slightly more abstract setup. A reader not interested in a more abstract treatment can simply skip the analysis below. I re-state the relevant insights and provide intuition in the context of the applications.

Suppose Alex's type is drawn from $\Theta \subset \mathbb{R}$ given a continuous prior $\pi_0(\theta)$. His preferences may again depend on his type, but the set of signals he observes do not. Alex's action y results in an outcome $x \in X \subset \mathbb{R}$ where X is finite. Tanya observes only x, but both Alex and Tanya might have private information about the state ω.

I impose two assumptions on the distribution of the outcome. First, the dependence of the outcome distribution on Alex's type and information can be sufficiently expressed by a set of conditional distribution functions

$$\pi(x \mid \theta, u^*(S)) : \Theta \times \mathbb{R} \to \Delta X$$

where $u^*(S)$ is Alex's expected payoff integrated over his type-dependent expected payoffs using the prior $\pi_0(\theta)$. Although this one-dimensional dependence of the outcome on Alex's information may seem restrictive, because observing more signals implies a higher expected utility for all types, monotone changes in information will affect u^* in a monotone manner as well. Furthermore, as Lemma 1 showed, information projection causes a monotone misperception of the value of Alex's information in maximizing his payoff. Second, the outcome process $\pi(x \mid \theta, u^*(S))$ satisfies the monotone-likelihood ratio property in θ and in the quantity u^*. The formal definition is adopted from Milgrom (1981).

Condition 1 *A set of densities $\pi(x \mid \theta, u^*(S))$ satisfies MLRP in θ if for all p and all $\theta > \theta'$ and $x > x'$,*

$$\frac{\pi(x \mid \theta)}{\pi(x' \mid \theta)} > \frac{\pi(x \mid \theta')}{\pi(x' \mid \theta')} \tag{3}$$

where for a fixed θ, $\pi(x \mid \theta) = E_p[\pi(x \mid \theta, u^*(S))]$. The analogous condition holds for all $u^* > u^{*\prime}$.

The next result shows that if the above assumptions are satisfied, Tanya's expected inference about Alex's type is decreasing in her bias. This result uses Lemma 1, which allows me to express the consequences of information projection in terms of Tanya's exaggeration of the distribution of u^*.

[6] In the Appendix, I show that this robust exaggeration of how much a projectee knows distinguishes my model from the earlier incomplete anchoring approaches adopted by CLW (1989) because anchoring-based approaches are generically inconsistent with this property of my model.

Theorem 1 *For all $\pi_0(\theta)$, $E_p[\pi_1^\rho(\theta)]$ decreases in ρ in the sense of fosd where expectations are taken with respect to the true distribution of signals. If $p = \widehat{p}$ except that $p^T \geq \widehat{p}^T$, then $E_p[\pi_1^\rho(\theta)]$ fosd $E_{\widehat{p}}[\pi_1^\rho(\theta)]$ for all $\rho > 0$.*

Since both more information and higher types shift the distribution of outcomes upwards, by exaggerating Alex's information, Tanya expects a higher outcome distribution than she should. This way, she is too surprised observing outcomes below any particular level. She always overweighs how much information low outcomes reveal and hence underweighs how much high outcomes reveal about Alex's type. As long as higher types increase the outcome distributions more, she underestimates Alex's type on average.

Above, I assumed that a higher outcome is good news about Alex's quality θ. Since the results depend only on the monotonicity assumption, when a higher outcome is bad news about θ – in the sense of Milgrom (1981) – a biased observer will overestimate θ on average, and when x is neutral about θ; no average misestimation is implied by information projection.

3.4 Substitute and Complement Information

Key comparative statics of the model will depend on how the value of the projected information relates to Alex's true information. I introduce an important distinction between complement and substitute information. I call two signals substitutes if the additional value of one decreases once the other is know. I call two signals complements if the opposite is true.

Suppose one needs to find a house in London. The street address and the full post code are substitute signals. In contrast, the latitude and the longitude coordinates of its location are complements. Only once both are know, does one have a reasonable chance of finding it.

Definition 2 *Given $u(\omega, y)$, two signals s_l and s_j are substitutes if $u^*(s_l \cup s_j) - u^*(s_j) < u^*(s_l) - u^*(\emptyset)$ and complements if $u^*(s_l \cup s_j) - u^*(s_j) > u^*(s_l) - u^*(\emptyset)$.* [7]

Finally, people may anticipate the bias of others. Let the probability density function $\psi_{i,k}(\rho) \in \Delta\,[0,1]$ describe the beliefs of person i about how much person $k \neq i$ projects information. If ψ_i is not concentrated on 0, person i believes that there is a non-zero probability that person k is biased.

4 Defensive Agency

A supervisor evaluates an agent whose task involves processing information. To fix ideas, a radiologist receives a noisy *X-ray* s_0, or set of radiographs S_0, on the condition of the patient ω and is asked to recommend a treatment y. A state- and action-dependent stochastic process $\pi(x \mid \omega, y)$, leads to a success x_S or a failure x_F.

[7] Since information need not be continous, the appropriate definitions are in terms of *super-* and *sub-modularity*. On the presence of non-concavities in the value of information see e.g., Radner and Stiglitz (1984).

Radiologists differ in skill. A radiologist's ability to interpret the $X-ray$ correctly depends on his type $\theta \in [0,1]$ distributed according to π_0. I assume that for any fixed set of ex-ante radiographs S_0, a higher type can identify the true state more often and that adding a signal to S_0 weakly increases this probability for all types.[8]

The evaluator observes the realization of x along with some novel information s_1 about ω. The evaluator uses all her information to learn about the radiologist' talent. Since observing an outcome, particularly whether a treatment failed or succeeded, typically provides plenty novel medical information about the patient, this assumption is both natural and very weak. Furthermore, since the outcome process is state-dependent, in the standard Bayesian setting, this helps the evaluator to form better estimates of θ. Let's denote the evaluator's ex-post assessment by π_1

To close the model, I assume that the radiologist prefers a higher reputation to a lower and thus maximizes an increasing function of the evaluator's posterior $b(\pi_1)$, where $b(\pi_1) \geq b(\pi_1')$ whenever π_1 fosd π_1'. If this is commonly known, then in the efficient Bayesian equilibrium of this game, the radiologist offers a recommendation y^* which maximizes the probability of success given his beliefs about the patient σ_1,

$$y^* \in \arg\max_{Y} E_{\sigma_1}[\pi(x_S \mid \omega, y)] \tag{4}$$

and given her beliefs about ω and the distribution of information, the evaluator updates her assessment of θ via Bayes rule.

4.1 Monitoring and Reputations

A Bayesian evaluator uses s_1 only to learn more about the ex-ante task. A biased evaluator also projects s_1 and exaggerates the ex-ante probability of success for all types. I call the proportional exaggeration of this probability for a specific type the value of the information gap for this type. Formally, given Eq.(4):

$$g_\theta = \frac{\pi(x_s \mid \theta, s_0, s_1)}{\pi(x_s \mid \theta, s_0)}$$

with g being the probabilistic average of g_θ given the prior on types. The first result is a direct application of Theorem 1 to binary outcomes.

Proposition 1 *For all π_0, expected reputation $E_x[\pi_1^\rho]$ is decreasing in ρ and in g in the sense of fosd.*

Since a biased evaluator exaggerates the ex-ante probability of success for all types, she overweighs the information revealed by a failure, and hence underestimates the radiologist's competence *on average*.

[8] A simple example is where θ is the probability that the radiologist correctly understands the content of s_0. Another is where it is the probability that the agent identifies the correct interpretation of different pieces of ex-ante evidence under binding time constraints. In devising a successful military response for example, the ability to sort information quickly is key.

The above result is true for all priors, hence long-run underestimation holds a fortiori. The next examples illustrate the comparative static results with respect to the value of the projected information, i.e., the information gap g.

In-depth Monitoring Obtaining additional signals about the true state of the world – the nature of the ex-ante task – leads to lower reputations on average. Knowing more about what happened during the collapse of Challenger allows investigators to determine the correct interpretation of the ex-ante evidence. More intensive monitoring translates into lower valuations on average.

Scapegoating after Scandals The above logic implies that launching an investigation against a person too easily turns this person into a scapegoat. The more the public learns about an alleged corruption case or a sex-scandal of a politician, the more likely it is that the public will think that she had poor judgement or character in the past. If one group of people (such as a salient minority group) is subject to more scrutiny than another, biased observers will ascribe worse qualities to the minority than to the majority.

Favoritism Similarly, information projection provides an endogenous mechanism for favoritism. A fair-minded supervisor has to rank workers based on cognitive skill and discipline alone. Her instructions will more easily be understood by a worker who shares the supervisor linguistic background than by a worker with a greater information gap such is an immigrant or a worker who speaks a different dialect.[9] By projecting information the supervisor misattributes the differences in linguistic backgrounds to differences in talent, and perceives the foreign worker as less focused and less talented.

While favoritism is typically explained as an exogenous preference of a supervisor for one worker over another, (Prendergast and Topel 1996), the model of information projection describes a different channel. Favoritism towards 'similar' types here arises endogenously as a function of the differences in information gaps. Furthermore, such favoritism is unintended.

The same logic implies that fair treatment can mistakenly appear as favoritism. It is commonplace for investigations into questionable police shooting and episodes in war to yield results in which internal investigations find no wrongdoing. The cops and military investigators presumably can appreciate the ex ante perspective better (i.e. they realize how difficult it could be to predict whether a person or enemy has a weapon or presents a threat, or is a combatant or civilian), but to hindsight-biased outsiders it appears exactly like favoritism.

4.2 Attributing Luck and Skill

I now turn to the impact of information projection on conditional rather than average assessments. Conditional assessments matters in understanding short and medium-run performance differences: whether evaluators attribute these to luck or skill. The results here depend on the type of the ex-post information revealed by monitoring. If knowing the projected ex-post information ex-ante would have helped higher types more than lower ones, evaluators misattribute differences in luck to skill and *over-infer* skill from performance. When the reverse is true, evaluators misattribute differences in skill to differences in luck.

[9] In Section 6 of the paper I derive a related example formally.

Proposition 2 *If g_θ is increasing in θ, then $\pi_1^\rho(\theta \mid x_S)$ is increasing in ρ in the sense of fosd. If g_θ is decreasing in θ, then $\pi_1^\rho(\theta \mid x_S)$ is decreasing in ρ in the sense of fosd.*

Over-inference In many professional settings, additional information helps higher types more. An additional radiograph helps a well-trained radiologist more than it helps a quack. A careful CEO can adjust a company's portfolio in anticipation of a future price-shock than someone who is negligent. If the price shock is truly an ex-post surprise however, the CEO's ex-ante effort makes little difference. By projecting information, board members will act as the CEO could have anticipated the shock and reward him for good luck and punish him for bad luck.

Football coaches are routinely fired after a bad season, but there is no clear evidence how much coaching skill actually matters. If principals project skill intensive ex-post information, they will exaggerate the role of talent in determining performance. Successful coaches, executives or politicians will be relatively over-valued and failed ones relatively under-valued.[10]

Under-inference In some professional settings, success itself is the creation of easy-to-understand information. This information may well have the same content as an ex-ante already existing piece, but can be successfully processed by a much broader set of types. The prime example is that of a simple mathematical proof or a commercial invention. By projecting information, observers here fail to recognize how hard the original problem was ex-ante, and underappreciate the skill-content of a successful solution.

A clear application is *patent* law. Legal professionals will fail to appreciate how non-obvious an invention was ex-ante and will grant too few patents. Mandel (2006) argues – in what he calls 'patently obvious" – that this is a significant problem in patent law and provides supportive experimental evidence. [11]

4.3 The Supply of Information

Let's now turn to the *key* result of the paper, the agency conflict caused by the fear of average underestimation. As mentioned before, evidence suggests that medical and legal professionals do fear biased evaluations. To understand the consequences of such a fear, let me turn to the optimal response of experts who anticipate biased evaluations.

Note first that no matter whether the radiologist anticipates projection or not, he will always want to offer the best diagnosis given his ex-ante information. This is true because his utility is always higher after a success than after a failure.

Lemma 2 *For all $\psi(\rho)$, the agent's best-response is to maximize the ex-ante probability of success.*

Anticipation of biased monitoring affects the radiologist's preference over which medical cases to undertake. The more valuable the ex-post medical information is, the worse reputa-

[10] Although the nature of the projected information is not controlled for, field evidence is consistent with systematic over-inference. Jenter and Kanaan (2011) find that boards do not filter out economic shocks from the evaluation of CEOs, and Wolfers (2007) provides similar evidence on the re-election of politicians following unanticipated price shocks.

[11] Here whether assessments after a failure here are too low or two high, depends on the two countervailing forces, under-estimation and under-inference.

tional prospect he faces. The best way to capture his strategic response is then to consider his preference over the ex-ante information contained in the case.

Suppose the radiologist has no direct choice over what medical cases to accept, but has discretion over what ex-ante information to order. In particular, he can decide whether to order, on top of s_0, an additional ex-ante radiograph, s_0'. The social value of producing s_0' is given by some value $\mathbf{a} \in \mathbb{R}$. Benefits of production include the additional knowledge gained, while costs include the alternative use of medical resources, the delay in treatment, and increased pain or radiation. The radiologist privately observes a and publicly decides to produce s_0' before taking action y_a. As standard in models of career concerns, e.g., Harris and Holmström (1982), I assume that when making this decision the radiologist faces the same uncertainty about his skill as the evaluator.[12]

To assume away all direct agency conflicts, I consider the case where the radiologist fully internalizes a and is risk-neutral over assessments π_1. I relax both assumptions in Section 5. The agent's augmented objective is now:

$$u(y, a, \omega) = \chi a + b(\pi_1^{\rho}(\theta)) \tag{5}$$

where χ equals 1 if s_0' is produced and 0 otherwise.

Let m be the ex-ante probability that the a full evaluation occurs after task completion. In the Bayesian case, the agent produces s_0' whenever it is socially optimal: there is a neutrality between the task solved and the expected reputation. In contrast when assessments are biased due to information projection, the same neutrality is violated. An agent best response is now to *over-produce* tests that are *substitutes* of, and *under-produce* tests that are *complements* of the information the evaluator independently learns ex-post.

Proposition 3 *For all π_0, s_0, s_1 the agent's best response is given by a cut-off strategy $a(m, \rho)$ where s_0' is produced if and only if $a \geq a(\rho, m)$. Furthermore,*
1. *For all ρ and m, $a(0, m) = a(\rho, 0) = 0$.*
2. *If s_0' and s_1 are substitutes, $a(\rho, m)$ is increasing in m and $\rho > 0$.*
3. *If s_0' and s_1 are complements, $a(\rho, m)$ is decreasing in m and $\rho > 0$.*

The above proposition implies that as the probability of monitoring increases, physicians will increase the production of substitute tests and decrease the production of complement tests. The common force driving these opposite responses is the desire to reduce the value of the information gap g between the ex-ante and the ex-post stages.

Consider the case of substitute information. Suppose the radiologist can order an additional MRI ex-ante and knows that independent of his production choice the content of this MRI will become available to evaluators ex-post. Whenever the marginal value of the ex-post information is lower once one knows the outcome of the ex-ante MRI, the radiologist has an additional incentive to produce this MRI. Physicians who fear information projection will take too many biopsies, expose patients to too much radiation, or recommend surgery

[12] The qualitative results of Proposition 3 will continue to hold when the amount of inference about θ is held constant and the radiologist has superior information about his own type and can be extended to reputational models with signalling.

too often so that their ex-ante predictions appear less incompetent in hindsight. Similarly, engineers might delay projects to collect information even if they know that the marginal value of this information is much lower than the additional cost of the delay.

The opposite happens when the result of an ex-ante test complements ex-post information. A physician who fears information projection avoids ordering complement information ex-ante. A noisy mammogram is often the best way to detect breast cancer at an early stage. Once the location of the tumor is known ex-post, an evaluator can easily determine whether a tumor was already developing ex-ante. Such ex-post insight is often impossible without the ex-ante mammogram. Ironically, the radiologist appears less incompetent on average if he does not order the mammogram ex-ante. In this fashion, doctors will be reluctant to order vague radiological charts, therapists will be reluctant to keep transcripts, social workers will be reluctant to make ambiguous phone calls, and engineers will be reluctant to keep equivocal forecasts of catastrophes, because in hindsight evaluators will interpret these as clean ex-ante evidence.

The above proposition implies that as the liability pressure on physicians increases it is the *composition* of the ex-ante information that changes. Whenever defensive practices cannot fully eliminate underestimation, the model makes the following joint prediction: as liability pressure rises, the reputation of the average physician drops and the medicine practiced becomes less efficient.[13] This joint prediction identifies the agency problem.

Proposition 3 is stated for the case where evaluators observe a binary outcome variable, but the same qualitative result holds with the generality of Theorem 1. This is true because if the conditions of Theorem 1 are satisfied, then from the fact that an increase in the information gap between the radiologist and the evaluator leads to a decrease in the expected reputational lottery – in the sense of first-order stochastic dominance – it follows that the radiologist wants to distort the production of information in the exact same ways. An additional ex-ante substitute signal increases and an additional complement signal decreases the reputational lottery.

4.4 Defensive Medicine

The agency conflict described above is consistent with evidence on defensive medicine described as procedures designed to minimize false liability rather than maximize cost-effective health care.Leonard Berlin's (2003) testimony to the US Senate Committee on Health, Education Labor and Pensions illustrates the potential scope of information projection in radiology. Based on empirical studies conducted in prestigious US medical institutions, he argues that in hindsight as many as many as 90% of lung cancers and 70% of breast cancers can be observed on radiographs previously read as normal.[14]

The study of Studdert et al. (2005) documents widespread defensive practices consistent with those described by Proposition 3. Studdert et al. (2005) interviewed physicians in

[13] The latter could – though need not – translate to both higher costs (substitute info) and lower quality (complement info).

[14] Berlin, L. (2003), *Statement of Leonard Berlin, M.D. To the U.S. Senate Committee on Health, Education Labor and Pensions Re: Mammography Quality Standards Act Reauthorization.* April 8, 2003. http://www.fda.gov/ohrms/dockets/ac/03/briefing/3945b1_05_Berlin%20testimony.pdf

high-risk fields in Pennsylvania. Of the 824 physicians interviewed, 93% engaged in defensive medicine involving both in 'assurance' and 'avoidance' behavior which correspond to the over- and under-production of medical tests. 90% of physicians reported ordering medically unwarranted diagnostic tests. The most common responses also included frequently avoiding the most efficient medical test. As an example, 36% of radiologists reported ordering unnecessary MRIs and 54% of radiologist reported that they often avoid ordering medically efficient mammograms.

Supporting the setup of this paper, in the economics literature, Kessler and Mclellan (1996) argue that the primary motivation for defensive medicine is not incomplete monetary insurance but the fear of reputational loss.[15] Consistent with the model's prediction, Kessler and Mclellan (1996) use data from elderly heart patients to show that reducing liability pressure significantly reduces medical expenditures without increasing mortality or medical complications. Kessler and McClellan (2002) also find that the main effect of cost reduction is on diagnostic rather than on therapeutic practices, which also supports the mechanism of information projection since there is likely to be room for greater information projection in diagnostic practices.

Defensive practices are sometimes attributed to physicians' fear of random judicial judgements. This statement alone has little explanatory or predictive power. If the cause of false liability however, is not random judgment, but fear that evaluators exaggerate the ex-ante accuracy of tests, as argued for example by Berlin and Hendrix (1998) or Berlin (2004), increased efficiency should operate through an observable and systematic change in the composition of the diagnostic practices as specified by the above result.

Proposition 3 can equivalently be expressed in the context of ex-ante choice of medical procedures or physicians' timing of medical tests. The under-production of complement tests then has an implication to the problem of *preventive* medicine. Physicians in my model will be reluctant to order early tests - especially if these will be accessible by evaluators ex-post – because in most scenarios looking at such radiographs once an illness has developed much later will allow evaluators to detect 'early signs' of the illness too easily.

5 Discussion and Repairs

Proposition 3 is the key behavioral prediction of the model. It describes an agency conflict which does not exist under Bayesian assumptions. In this Section, I first demonstrate the robustness of this result and argue that instruments which correct for agency conflicts that arise due to agent's risk-aversion or because agents do not internalize the costs or benefits of information production are either ineffective or exacerbate defensive practices. I then turn to possible repairs and show how particular performance-sampling rules can mitigate defensive practices while not distorting Bayesian incentives.

[15] Kessler and McClellan (1996) argue that since virtually all physicians in their sample are fully insured against the financial costs of malpractice (damages, legal expenses etc.) physicians "may employ costly precautionary treatments in order to avoid nonfinancial penalties such as fear of reputational harm, decreased self-esteem from adverse publicity" etc.

5.1 Extensions

Let me now demonstrate the robustness of Proposition 3 by allowing for (i) agents who are risk-averse, (ii) agents who do not fully internalize the cost or benefit of production, and (iii) production where s_0' can be hidden from the evaluator.

Risk Aversion So far, I assumed that the radiologist was risk-neutral over assessments. The same qualitative results hold under risk-aversion. This is true because information projection affects the mean rather than the spread of the reputational lottery. In fact, risk-aversion typically amplifies defensive practices. As noted before, (Holmström, 1982), under correct expectations, risk averse agents have a preference for limiting the production of all skill-intensive information to prevent evaluators updating their beliefs.[16] In contrast, a risk-averse radiologist who fears underestimation will display *differential* responses. She would want to over-produce skill-intensive substitute information. Hence, production inefficiency arises independent of the problem of optimal risk-sharing, the same directional effects hold even when the amount of inference about the radiologist's skill – the spread of the reputational lottery – is held constant.

Incentives A moral hazard problem arises when radiologists do not fully internalize the costs or the benefits of producing s_0' and calls for stronger or weaker production incentives. A tax on production or a bonus after a success can correct for distortions in the volume, but not in the composition of the information produced ex-ante. A radiologist who is monitored by biased evaluators still enjoys a higher reputation if he distorts production in the manner described above.

Similarly, even if a third-party who was aware of the evaluator's bias, such as a principal, inflated or deflated the evaluator's assessments, this would not limit defensive practices. Suppose the principal decided to inflate the evaluator's ex-post assessment π_1^ρ by shifting this distribution 'upwards' in the sense of first-order dominance (for example such that in expectations it equaled π_0). Since by distorting the ex-ante production of information, the radiologist can still affect the level of π_1^ρ in the same way as before, his defensive incentives are unchanged. For any fixed inflation or deflation policy, the radiologist would still want to distort the ex-ante composition of tests as described by Proposition 3. Differential incentives for the production of substitute and complement information would be required to reduce defensive medicine.

Observability Finally**,** I also assumed above that if a signal was produced ex-ante, the evaluator observed this. If the expert could produce s_0' secretly, that is, in a way that the evaluator believed s_0' was *not* produced, then the expert would want to produce this signal too often, independent of whether it is a substitute or a complement of s_1. Such over-production is not specific to the biased case. Secret production of information has the additional benefit of raising reputations even when evaluators do not project information: it fools the evaluator into believing that the expert has solved a harder problem than he actually did.

[16] On this point see also Hermalin (1993).

5.2 Performance Sampling

Under Bayesian assumptions, ex-post information helps the evaluator form better estimates of θ. In the biased case, ex-post information leads to biased beliefs and a costly agency conflict. Maintaining observability but suppressing ex-post information from monitoring is typically not a viable strategy. The observation of an outcome typically cannot be separated from the observation of novel information about the task. The only possibility is to selectively suppress outcome information.

Consider the same setup as before, but emphasize more explicitly that the agent completes a sequence of independent tasks. Evaluators then need to sample from this sequence. Furthermore, the sample need not be balanced: successful cases may knowingly be over-sampled and failed ones under-sampled. For example, in deciding on the promotion of a police officer, his superiors might devote more careful attention to cases where the officer identified the offender quickly than to cases where his efforts produced no timely results. In academia or industry research, promotion committees might only sample projects that led to decent inventions and have no detailed access to projects where the candidate ended up wasting resources.

To be more specific, let me define a performance sampling rule. The definition below is clearly a limit case of a more realistic scenario where observability varies more smoothly.

Definition 3 *A sampling rule is a probability pair (m_f, m_s) such that given a task the evaluator observes (S_0,s_1,x) with probability m_f if the outcome is a failure and probability m_s if the outcome is a success. In the remaining cases, the evaluator only observes the structure of S_0.*

A performance sample allows for meaningful inference if the evaluator knows the sampling rule. Suppose that the sampling rule is common knowledge. Evaluators can then always infer the true success/failure ratio from the observed sample. Hence, in the Bayesian case the specification of the sampling rule has no effect on expected reputations. Agents here have no reason to distort information production.

Lemma 3 *If $\rho = 0$, then $a(m, 0) = 0$ for all m_s,m_f.*

Consider now the biased case. The fact that (m_f, m_s) is common knowledge no longer guarantees efficient production. Instead, the specification of the sampling rule directly affects production. This follow from the fact that observers receive relevant ex-post information only if a case is sampled. By selecting a sampling rule (m_f, m_s) a designer can influence whether evaluators are exposed to ex-post information more often after a failure or after a success. Since conditional assessments are not necessarily biased in the same direction after a success and a failure, an increase in m_s versus an increase in m_f might affect defensive practices in different ways.[17]

[17] A separate issue from the one discussed in the text is when evaluators are fooled about the sample. In the agency setup fooling will not solve the problem. Suppose evaluators have wrong beliefs about (m_s, m_f). This would obviously affect the level of the radiologist's reputations, but as argued in the beginning of this Section, such inflation or deflation of reputations will not affect the incentives for defensive practices qualitatively. It

The analysis of how the exact choice of the sampling rule affects production is complicated by the presence of over- and under-inference as described by Proposition 2. To simplify matters, let's first focus on the case where these forces are not present and thus information projection leads only to underestimation. Assume also that simply inferring the outcome of a case provides no additional information about the state. Then investigating the reasons for successful performance but never for failures will eliminate the incentives for defensive practices entirely.

Proposition 4 *Suppose g_θ is constant in θ for all s_0. The cut-off $a(m,\rho)$ is constant in m_s.*

1. *If s_0' and s_1 are substitutes, $a(\rho,m)$ is decreasing in m_f and $\rho > 0$.*
2. *If s_0' and s_1 are complements, $a(\rho,m)$ is increasing in m_f and $\rho > 0$.*
3. *For all $\rho > 0$, $a(m,\rho) = 0$ if and only if $m_s \geq 0$ and $m_f = 0$.*

Suppose that by projecting information evaluators exaggerate each type's productivity by the same factor. Information projection here only distorts assessments after a failure – Proposition 2. If evaluators only sample cases where the expert succeeded, they only observe the ex-post information after a success. Of course from the number of successful cases she observes, the evaluator *infers* the radiologist's rate of failure, but when thinking about those failed cases she is constraint to adopt the ex-ante perspective. This eliminates the distortion due to information projection and guarantees that the ex-ante expected reputation and production incentives of the expert are unbiased.

The same qualitative insight holds more generally. A sufficient condition can be expressed as a general complementarity relation between skill and ex-post information: the greater is the average additional value of the ex-post information the more it increases the productivity of high types relative to low types. Under such conditions the incentives for defensive practices are again best reduced when the probability that a successful case is reviewed is higher than the probability that a failed case is reviewed. Let g'_θ denote the type-dependent information gap when s_0' is also produced and g_θ denote the same when s_0' is not produced.

Proposition 5 *Suppose g_θ is constant and $(g' - g)\frac{\partial g'_\theta}{\theta} > 0$. Then for all $\rho > 0$, $a(m,\rho)$ is decreasing in m_s and increasing in m_f and there exists $m_s^*(\rho) > m_f^*(\rho) > 0$ such that $a(m^*(\rho),\rho) = 0$.*

Organizational Repairs. Selectively suppressing information can achieve three goals simultaneously: (i) allow for some inference about the quality of the expert, (ii) leave Bayesian production incentives unaffected, and (iii) improve production efficiency in the biased case. Doing so involves a departure from organizational practices that would be optimal under Bayesian assumptions. Consider first the case where resource constrains on monitoring are non-binding. Here, in the unbiased case, a promotion committee would want to review all medical cases. In the presence of information projection – or experts' fear thereof – a promotion committee shall instead review only a limited sample. The focus of

would simply shift $a(m,\rho)$ up or down, but will not change the comparative statics.

monitoring could also change as organizations become more concerned about information projection. Suppose resource constraints on monitoring are binding. Learning about the causes of failures could be more informative under Bayesian assumptions because this may allow for a better separation of low and high types. The above results imply that the presence of information projection may shift the optimal allocation of monitoring resources from investigating failures to investigating successes. Future research can address these issues in more detail.

6 Communication

I now turn to a brief application of the model to the problem of communication. The main prediction here is that privately informed speakers will assume too much knowledge on behalf of their listeners. A biased speaker will fail to tailor her message to the actual background knowledge of her listener, and upon observing that as a result the listener acts more 'confused', she will infer that he is dumb, inattentive or unable to execute.

Computer manuals or interfaces that are ideal for engineers are too technical for the average user. The value of a message is determined not only by its content, but also by its relation to what the audience knows to begin with. Describing information in a way which builds on the existing knowledge of listeners is key for efficient communication.

To understand the impact of projection bias on communication in the presence of such complementarity, I focus on the simplest coordination problem where a sender and a receiver have perfectly aligned interests. The sender sends a message y_s, the receiver responds with an action y_r. The receiver's objective is to match the true state ω as often as possible. Suppose the true state ω is decomposable into $\varpi_1, \varpi_2 \in \{-1, 1\}$ the following way

$$\omega = \varpi_1 * \varpi_2$$

There are three signals: $s_1 = \varpi_1$ the '*technical language*', $s_2 = \varpi_2$ the '*technical term*', and s_3 a noisy '*lay description*'.The sender knows both s_1 and s_2. These two pieces of information are perfect complements: one conveys information if and only if the other is also known. The lay description is an imperfect substitute of these two, and is given by s_3 such that $\Pr(s_3 = \omega \mid \omega) = h < 1$.

The notion of a technical term is simple: the correct interpretation of most medical, technological or mathematical terms requires extensive prior knowledge of the relevant field. To illustrate how this differs from a lay description, suppose communication takes place under binding time constraints. It takes much more time to effectively convey the meaning of a medical diagnosis or an engineering idea to a layperson than to a trained doctor or engineer. Hence, when time is binding, the lay description ultimately conveys less information and thus can be modeled as less precise (needing to ignore important details), and hence allows the listener to identify the truth less often.[18]

[18] Alternatively, the results below can be re-phrased as a speaker's decision on how much time to allocate to communication.

Suppose that communicating the technical language is prohibitively time-consuming. The sender thus faces three options: (i) send s_2, (ii) send s_3, or (iii) remain silent. The cost of sending a message is c, and silence is free.[19] Suppose that there is a symmetric prior on ϖ_1 and ϖ_2.The table below summarizes a ρ-biased speaker's perception of the probability of a correct choice by the receiver for each of the messages:

silence	technical term	lay term
$\frac{1}{2}+\frac{\rho^2}{2}$	$\frac{1}{2}+\frac{\rho}{2}$	$h+(1-h)\rho^2$

The proposition below summarizes a biased person's communication strategy. Let $\rho^* = (h-\frac{1}{2})/(1-h)$.

Proposition 6 *If $\rho < \rho^*$, the speaker sends s_3 if $(1-\rho^2)(h-\frac{1}{2}) > c$, and stays silent otherwise. If $\rho > \rho^*$, the advisor sends s_2 if $\frac{1}{2}(\rho-\rho^2) > c$, and stays silent otherwise.*

Consistent with earlier intuitions, biased experts send messages that are too difficult for lay people to understand. Relative to the unbiased case, information projection distorts communicational efficiency in two specific ways: experts speak too rarely, and when they speak, they send messages that are too technical. The above proposition shows that overall a biased speaker under-communicates information that substitutes for her expertise (s_3) and over-communicates information that complements it (s_2).

It is easy to see, following Theorem 1, that if the speaker makes inferences about the attentiveness or his ability to utilize information by observing y_r, she will be too surprise how often the listener's action misses the true state. The above result illustrates that the option to send a technical term, because it wrongly increases the speaker's belief of how much information can be transmitted, exacerbates such underestimation.

This above result is consistent with the evidence on the misattribution of the causes of communication failures, as mentioned in Section 2. The following story collected by Cusumano and Selby (1995) illustrates the above mechanism:

> At one point, Microsoft started surveying users to see how many of them found it easy to use a particular feature. Software developers often refused to believe the statistics. "The usability group would tell the development group 'Six out of ten couldn't do this.' And the developer's reaction would be, 'Where'd you find six dumb people?" The usual nonsense answer 'Well, they can just look in the manual if they don't know how to use it,' or 'My idea is brilliant; you just found ten stupid people'.

Favoritism and Protocols The presence of information projection provides a rationale why communication protocols can improve efficiency. Importantly though, even in our simple example, mandating communication need not restore efficiency. In fact, when $\rho > \rho^*$, such a policy will backfire. Here banning silence will only increase experts' negative attributions about listeners. Instead, protocols that restrict the use of messages which complement the

[19] To simplify the analysis, I assume that the speaker knows the properties of s_3, but not its realization. The insights easily extend to the case where this assumption is relaxed.

speaker's background, but encourage the use of messages which substitute for it are necessary to restore efficient information transmission.

A close analogue of communication protocols is *proof-reading*. The above result implies that the best people to judge the appropriateness of a text-book or a doctor-patient communication-protocol are not fellow experts, but people with less expertise.

Communication protocols can also help reduce favoritism in the workplace. Since in their interactions with employees supervisors misattribute differences in information to differences in skill, workers such as immigrants who do not share the linguistic background of the supervisor will suffer relative to workers who do. This might prompt skilled immigrants to choose less productive self-employment, such as opening a restaurant, rather than joining a corporate hierarchy. [20]

7 Discussion and Conclusion

The presence of hindsight bias, curse-of-knowledge and the illusion of transparency in human judgement illustrates a general mistake in information processing in the interpersonal context. I termed this mistake information projection. The aim of this paper has been to enrich the economic analysis of situations with asymmetric information by introducing a widely-applicable model of this information projection capturing a broad class of mispredictions in a parsimonious manner. Applications to agency and communication settings demonstrated the relevance of the model for economics.

The key application of the model introduced a novel form of agency conflict which does not exist under Bayesian assumptions. In this context, the model provided a unified explanation of the types of assurance and avoidance behavior that medical observers have long attributed to defensive practices rather than cost-effective health care. To protect their reputation, fully concerned and risk-neutral workers under-produced ex-ante tests that complemented the information evaluators independently learned ex-post, and over-produced skill-intensive ex-ante tests that substituted for the same information.

Various tests can help identify the model. The results on belief-updating allow for tests in dynamic inference problems. Information projection can be tested on disaggregated choice data such as combining choices over information sets and choices over outcomes, (e.g., Camerer, 1992 or Loewenstein et al., 2006). Such designs also allow to further understand the extents to which people are sophisticated about their own tendency as well as the tendency of others to project information. Non-standard comparative static results of the paper, such as the impact of monitoring on the value of agency relationships, can help identify the presence and significance of information projection in economic data more generally.

An important limitation of my approach should be noted when testing the model. As many economic models, my model allows for heterogeneity in information projection, but it does not pin down such heterogeneity. Certain pieces of information can and are projected to a greater extent than others, but the model is limited by the fact that without additional

[20] Madarasz (2009) briefly discusses the affect of information projection on listener's understaning of messages, and argues that biased listeners fail to rely on their own private information sufficiently and hence will follow *herds* too often.

data, it cannot predict heterogeneity in a systematic manner. The key implications of the model and the results do not depend on the nature of such heterogeneity, but the model here provides a framework in which such heterogeneity can be established empirically and where various more ad hoc measures of informational mispredictions can be integrated.[21]

Information projection is likely to matter in agency problems and organizational settings not covered in this paper. Madarász (2009) contains an example where the model is applied to a problem with moral hazard. In a setting with limited liability and explicit contracting, information projection transforms a de jure *negligence rule* to de facto *strict liability*. Finer monitoring and steeper incentives can backfire and induce physicians to exert less care. Optimal monitoring is coarser and incentives are weaker than under Bayesian assumptions. Empirical applications of the model could also help us understand anomalous evidence on CEO turnover, (Jenter and Kanaan, 2011), politician turnover (Wolfers, 2007), or unintended favoritism. Future work can also address optimal contracting problems in the presence of biased evaluations more generally.

The model applies to various other social inference problems. The dinner example of Section 3 can directly be extended into a more powerful analysis of social conflict. People in many natural learning environments people will mistakenly attribute taste differences to differences in social intentions, leading to segregation and false perceptions of hostility. Interventions that re-shape the structure of learning could fix broken links, and greatly reduce false conflict. Further lab and field evidence could also help test these predictions.

Although the paper focused on interpersonal information projection, it is possible to extend the model to the intrapersonal domain. For example, when considering a person's own future selves, the model predicts that people will be overconfident about their prospective memory for those pieces of information that they currently know, but less so for information they know they will learn in the future. Exploring this link can shed new light on various puzzles in intertemporal choice, and provide novel predictions on the role of deadlines and reminders in mitigating self-control problems.

Finally, one possible theory of the source of information projection is limited perspective-taking, as discussed by Piaget and Inhelder (1967). In line with this interpretation, one can extend the model to consider the problem of *ignorance projection*, where people underestimate the probability with which signals whose realizations they do not know, are unavailable to others.[22] In the context of Bayesian games, Madarász (2011) develops a unified model of information and ignorance projection.

8 Appendix

Below, I first demonstrated that the incomplete model of anchored expectations employed by CLW (1989) generically violates information projection. I then turn to the proofs. Distributions and random variables below depend on the underlying state ω. To save on notation

[21] See e.g., Pohl et al (1993).

[22] Ignorance projection is unlikely to play a significant role in the contexts studied in this paper but it might be relevant in other economic contexts. The experimental paradigms of Fischhoff (1975), Newton (1990) and CLW (1989) do not allow to test for ignorance projection.

whenever ω is suppressed, I mean the expected distribution of the relevant variable after taking the expectations with respect to $\sigma(\omega)$.

Projection versus Anchored Expectations CLW (1989) offer an anchoring-based explanation of the curse of knowledge. There a strictly better informed Tanya perceives the *mean* expectation of a strictly lesser informed Alex to be the convex combination of her mean expectation and Alex's true mean expectation. This approach leaves other moments of Tanya's belief unspecified and does not address the case where information is not strictly ordered.

Let me now briefly show that generically the two approaches contradict each other. This fact is based on the observation that the exaggeration of the *proximity of two means* is not a measure of informational closeness. Thus anchored expectations imply neither that Tanya overestimates nor that she underestimates of the value of Alex's information. No-matter how one completed this anchoring-based account, anchored expectations violate information projection and vice versa.

Consider a numerical example, where Let there are three people with strictly ordered information about the return on an asset. 1. *Alex* is uninformed and has a uniform prior over $[0, 1]$. 2. *Gremin* is better informed and receives valuable information by learning that the return is either 0 or $\frac{3}{4}$ with equal probability. 3. *Tanya* learns that the true return is $\frac{3}{4}$. The distance between *Tanya's* and the least informed *Alex's* mean belief is $\frac{1}{4}$. The distance between the *Gremin's* and *Tanya's* mean beliefs is larger $\frac{3}{8}$.

Proof of Lemma 1. Let the belief over Ω induced by the information set S^i be σ_i. Analogously, let the belief induced by $S^i \cup S^k$ be σ_{i+k}. Since both Ω and Z are finite, σ_i and σ_{i+k} are also finite. We can collect the realizations of these posteriors as a function of the state-dependent signal realizations into matrices Σ_i and Σ_{i+k} respectively. By the law of iterated expectations, $E[\sigma_{i+k} \mid S^i] = \sigma_i$ for all I_i and I_k. This implies that there exists a non-negative matrix with columns summing up to 1, i.e., a Markov matrix, T such that $\Sigma_i = T\Sigma_{i+k}$. From the classic theorem of Blackwell (1953) on comparison of information sets it follows that for any fixed von Neumann - Morgenstern utility function, $u_i(y, \omega)$, given a finite action set Y, $\max_{y_i} u_i(y_i, \omega \mid S^i) \leq \max_{y_i} u_i(y_i, \omega \mid S^k \cup S^i)$.

Person k's Bayesian and fully biased perception of $i's$ information are given by p_i^0 and p_i^1. Both of these are probability distributions over $\mathbb{N}$, and one can obtain p_i^1 from p_i^0 by allocating probabilistic weights from Blackwell less informative to Balckwell more informative signal sets, given the above observation. It then follows from the above observation that $\pi^1(u_i^*)$ first-order stochastically dominates $\pi^0(u_i^*)$. Since $\pi^\rho(u_i^*)$ is a probabilistic mixture of $\pi^1(u_i^*)$ and $\pi^0(u_i^*)$, the same is true whenever comparing $\rho' > \rho$. Finally, projecting a more informative information set S' instead of S means shifting probabilistic weight to information sets with even greater expected utility. .

Proof of Theorem 1. Let X denote the cardinality of the outcome space and let us index the elements of X by l in ascending order. Below, I suppress the term u^* in the notation because it appears in all relevant terms.

By the law of conditional probability when $\rho = 0$, $E_x[\pi_1^0(\theta \mid x)] = \pi_0(\theta)$. Consider the case when $\rho > 0$, the expected belief of a ρ-biased observer that $\theta = \widehat{\theta}$ is:

$$E_x\pi_1^\rho(\widehat{\theta}) = \sum_{l=1}^{X}[(\frac{\pi^\rho(x_l \mid \widehat{\theta})\pi_0(\widehat{\theta})}{E_\theta\pi^\rho(x_l)})E_\theta\pi(x_l)] \tag{6}$$

which one can re-write as:

$$E_x\pi_1^\rho(\widehat{\theta}) = \pi_0(\widehat{\theta})\overbrace{\sum_{l=1}^{X}\pi^\rho(x_l \mid \widehat{\theta})\frac{E_\theta\pi(x_l)}{E_\theta\pi^\rho(x_l)}}^{\lambda^\rho(\widehat{\theta})} \tag{7}$$

The term $\lambda^\rho(\widehat{\theta})$ expresses the weight attached to type $\widehat{\theta}$. In the Bayesian case, $\lambda(\widehat{\theta}) = 1$ for all $\widehat{\theta}$, a result known as the martingale property of Bayesian beliefs. Below I show that if $\rho > 0$, the function $\lambda^\rho(\widehat{\theta})$ is decreasing in $\widehat{\theta}$. Hence low values of θ are overweighted relative to high values which implies that the biased posterior is below the Bayesian posterior in the sense of fosd.

1. Given the MLRP of $\pi(x \mid \theta)$ in θ, it follows that for any $l < X$ and $\widehat{\theta} > \theta'$, $\sum_{l=1}^{L} f[\pi^\rho(x_l \mid \widehat{\theta}) - \pi^\rho(x_l \mid \theta')] < 0$.

2. Given the MLRP of $\pi(x \mid \theta)$ in u, $E_\theta\pi(x_l)/E_\theta\pi^\rho(x_l)$ is decreasing in l for any $\rho > 0$.

3. Combining the above two facts, for any $L < X$, there exists $Z > 1$ such that

$$\sum_{l=1}^{L}[\pi^\rho(x_l \mid \widehat{\theta}) - \pi^\rho(x_l \mid \theta')]\frac{E_\theta\pi(x_l)}{E_\theta\pi^\rho(x_l)} = Z\sum_{l=1}^{L}[\pi^\rho(x_l \mid \widehat{\theta}) - \pi^\rho(x_l \mid \theta')] < 0 \tag{8}$$

4. Using the same logic for x_X, we have that $\pi^\rho(x_X \mid \widehat{\theta}) - \pi^\rho(x_X \mid \theta') \geq 0$ and that $E_\theta\pi(x_X)/E_\theta\pi^\rho(x_X) < 1$.

5. Given that $Z > 1 > E_\theta\pi(x_X)/E_\theta\pi^\rho(x_X)$ it follows that

$$\lambda^\rho(\widehat{\theta}) - \lambda^\rho(\theta') = \sum_{l=1}^{X}[\pi^\rho(x_l \mid \widehat{\theta}) - \pi^\rho(x_l \mid \theta')]\frac{E_\theta\pi(x_l)}{E_\theta\pi^\rho(x_l)} < 0 \tag{9}$$

   because $\sum_{l=1}^{X}\pi^\rho(x_l \mid \theta) = 1$ for any θ.

6. Given that $\lambda^\rho(\theta)$ is decreasing in θ, for any prior $\pi_0(\theta) \in \Delta\Theta$ and any θ^* it follows that:

$$\int_{\theta<\theta^*}\pi(\theta)\lambda^\rho(\theta)d\theta \geq \int_{\theta<\theta^*}\pi(\theta)\lambda^0(\theta)d\theta \tag{10}$$

   thus $E_x\pi^0(\theta \mid x)$ first order stochastically dominates $E_x\pi^\rho(\theta \mid x)$.

7. To show that the same relation holds for $\rho < \rho'$ note that $\pi^{\rho'}(x_l)\pi^\rho(x_{l'}) > \pi^\rho(x_l)\pi^{\rho'}(x_{l'})$ whenever $x_l > x_{l'}$. Combined with the fact that $\pi(x \mid \theta)$ satisfies the MLRP in u the proposition follows.

8. To show that the same relation holds for $g < g'$ note that the same relation hold as in the comparison of $\rho < \rho'$.

.

Proof of Proposition 1. Note that $\pi(x_S \mid S^0, \theta)$ satisfies the MLRP in θ and in signal additions to S^0 for any S^0. This is true because the probability of success is increasing in type and in signal addition. Thus we can apply Theorem 1 for the case where $X = 2$. .

Proof of Proposition 2. Suppose g_θ is increasing in θ. I first show that for all $\theta^* < 1$ the following inequality is satisfied:

$$\frac{E_\theta \pi^\rho(x_S)}{E_\theta \pi(x_S)} \geq \frac{\int_0^{\theta^*} \pi^\rho(x_S \mid \theta)\pi_0(\theta)d\theta}{\int_0^{\theta^*} \pi^0(x_S \mid \theta)\pi_0(\theta)d\theta} \tag{11}$$

This is true because if for some $\theta^* < 1$ the inequality is violated, $\pi^\rho(x_S \mid \theta^*)/\pi^0(x_S \mid \theta^*)$ would not be increasing contradicting our assumption. Using the fact that for any $\widehat{\theta}$,

$$\pi_1^\rho(\widehat{\theta} \mid x_S) = \pi_0(\widehat{\theta})\frac{\pi^\rho(x_S \mid \widehat{\theta})}{E_\theta \pi^\rho(x_S)}$$

we can re-arrange Ineq. (11) and get that

$$\int_0^{\theta^*} \pi_1^\rho(\theta \mid x_S)d\theta \leq \int_0^{\theta^*} \pi_1^0(\theta \mid x_S)d\theta \tag{12}$$

The proof for the case when g_θ is decreasing in θ is analogous. Finally, if g_θ is constant, $\pi_1^\rho(\theta \mid x_S) = \pi_1^0(\theta \mid x_S)$ for all θ and Bayesian and biased conditional beliefs after a success are identical. .

Proof of Proposition 3. To determine the incentives for production, we compare the information gaps in the cases where s_0' is produced and where it is not. Let the respective quantities be g' and g. I first show that there is additional incentive to produce s_0' iff $g' < g$. By setting $a = 0$, the agent's decision to produce s_0' is determined by whether the following holds or not:

$$b(\pi_1^\rho(s_0')) - b(\pi_1^\rho) \geq 0$$

This inequality holds if $\pi_1^\rho(s_0')$ fosd $\pi_1^\rho(\theta)$ which is true if $g' < g$.

If s_0' and s_1 are substitutes, then $u^*(s_0' \cup s_1) - u^*(s_0') < u^*(s_1) - u^*(\emptyset)$ which implies that

$$\frac{u^*(s_0' \cup s_1)}{u^*(s_0')} - 1 < \frac{u^*(s_1)}{u^*(s_0')} - \frac{u^*(\emptyset)}{u^*(s_0')} < \frac{u^*(s_1)}{u^*(\emptyset)} - \frac{u^*(\emptyset)}{u^*(\emptyset)}$$

and hence $g' < g$. Similarly, if s_0' and s_1 are complements, then $g' > g$. Thus the production incentives are strictly positive if $\rho > 0$ and $g' > g$, and strictly negative if $g' < g$.

By the law of iterated expectations, $a(0,0) = 0$ and by continuity, $a(m,\rho)$ is decreasing in m if $g' < g$ and increasing in m if $g' > g$. Similarly since $|g - g'|$ is increasing in ρ, $a(m,\rho)$ is increasing in ρ if s_0' and s_1 are complements and decreasing in ρ if they are substitutes. .

Proof of Lemma 2. Follows from the law of iterated expectations. .

Proof of Proposition 4. I suppress s_0 in the notation below. The expected ρ-biased posterior depends both on the success probability and the monitoring frequencies. Given a fixed (m_f, m_s), if s_0' is not produced, we have that $E_x\pi_1^\rho(\theta)$ equals

$$E_\theta\pi(x_S)[m_S\pi_1^\rho(\theta \mid x_S) + (1-m_S)\pi_1(\theta \mid x_S)] +$$
$$(1 - E_\theta\pi(x_S))[m_F\pi_1^\rho(\theta \mid x_F) + (1-m_F)\pi_1(\theta \mid x_F)]$$

An analogous expression holds where s_0' is also produced.

If g_θ is constant, $\pi_1^\rho(\theta \mid x_S, s_1) = \pi_1(\theta \mid x_S)$ and hence $m_F = 0$ implies that ρ-biased expected assessments equal the prior, i.e., $a(m,\rho)$ is constant in m_S. Since $\pi_1^\rho(\theta \mid x_F)$ is decreasing in g, $a(m,\rho)$ is increasing in m_F if $g' > g$ and decreasing if $g' < g$. As a consequence, $a(m,\rho) = 0$ for all $\rho > 0$ iff $m_S \geq 0$ and $m_F = 0$. .

Proof of Proposition 5. We need to specify conditions on (m_S, m_F) such that expected assessments are independent of production. For the expected assessments across the two production scenarios to equal, the following condition must be met:

$$m_F[\pi_1^\rho(\theta) - \pi_1^\rho(\theta \mid s_0')] = E_\theta\pi(x_S \mid s_0')[m_S - m_F][\pi_1^\rho(\theta \mid x_S, s_0') - \pi_1(\theta \mid x_S, s_0')]$$

To derive the above result, I used the fact that if s_0' is not produced, $\pi_1^\rho(\theta \mid x_S) = \pi_1(\theta \mid x_S)$ for all ρ. From the fact that $(g' - g)\frac{\partial g_\theta'}{\theta} > 0$ it follows that the LHS and the RHS of the above equation have the same signs iff $m_S > m_F$. Furthermore, if $m_S > m_F$, then $m_F/(m_S - m_F) \in [0,\infty)$ and there always exist $m_S^*(\rho) > m_F^*(\rho)$ such that the above equality holds. The comparative statics follow from the earlier discussions. .

Acknowledgement 2 *This paper is based on Chapter 2 of my dissertation submitted to the University of California at Berkeley. I would like to thank the anonymous referees and the Editor Bruno Biais for their very helpful suggestions. I also thank George Akerlof, Jean-Pierre Benoit, Colin Camerer, Gary Charness, Vince Crawford, Marina Halac, Daniel Kahneman, Botond Kőszegi, Ulrike Malmendier, Andrea Prat, Marit Rehavi, Joel Sobel, Adam Szeidl, various seminar participants, and in particular Matthew Rabin for helpful comments. All errors are mine.*

References

[1] Anderson, J., M. Jennings, J. Lowe and P. Reckers. (1997), "The Mitigation of Hindsight Bias in Judges' Evaluation of Auditor Decisions." *Auditing: A Journal of Practice and Theory*, Vol. 16, 20–39.

[2] Arkes, H., P. Saville, R Wortman and A. Harkness. (1981), "Hindsight Bias Among Physicians Weighing The Likelihood of Diagnoses." *Journal of Applied Psychology*, Vol. 66, 252 – 254.

[3] Baron, J. and J. Hershey. (1988), "Outcome Bias in Decision Evaluation." *Journal of Personality and Social Psychology*, Vol. 54, 569-579.

[4] Biais, B. and M. Weber. (2009), "Hindsight bias and Investment Performance." *Management Science,* Vol. 55, 1018-1029.

[5] Berlin, L. and R. Hendrix. (1998), "Perceptual Errors and Negligence." *American Journal of Roentgenology*, Vol. 170, 863-867.

[6] Berlin, L. (2004), "Malpractice Issues in Radiology: Outcome Bias." *American Journal of Roentgenology*, Vol.183, 557-560.

[7] Blackwell, D. (1953), "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics*, Vol. 24, 265-272.

[8] Camerer, C., G. Loewenstein and M. Weber. (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy,* Vol. 97, 1234-1254.

[9] Camerer, C. (1992), "The Rationality of Prices and Volume in Experimental Markets." *Organizational Behavior and Human Decision Processes*. Vol. 51, 237-272.

[10] Camerer, C., Ho T., C. Juin-Kuan. (2004), "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics.* Vol. 119, 861-896.

[11] Caplan, R., K. Posner and F. Cheney. (1991), "Effect of Outcome on Physician Judgments of Appropriateness of Care." *Journal of the American Medical Association,* Vol. 265, 1957-1960.

[12] Cusumano, M. and R. Selby. (1995), *Microsoft secrets.* New York: Free Press.

[13] Crawford, V. M. Costa-Gomes and N. Iriberri. (2010), "Strategic Thinking." *mimeo* Oxford.

[14] Durell, A. (1999), "Attribution in Performance Evaluation." Ph.D Thesis. Harvard University, Cambridge MA.

[15] Eyster, E. and M. Rabin. (2005), "Cursed Equilibrium." *Econometrica*, Vol. 73, No. 5., 1623-1672.

[16] Fischhoff, B. (1975), "Hindsight / foresight: The Effect of Outcome Knowledge On Judgement Under Uncertainty." *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 1, 288-299.

[17] Gilovich, T., K. Savitsky and V. Medvec. (1998), "The Illusion of Transparency: Biased Assessment of Other's Ability to Read our Emotional States." *Journal of Personality and Social Psychology*, Vol. 76, 743-753.

[18] Gilovich T., V. Medvec, and K. Savitsky. (2000), "The Spotlight Effect in Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance." *Journal of Personality and Social Psychology,* Vol. 78, 211-222.

[19] Guilbault, R., F. Bryant, J. Brockway, and E. Posavac. (2004), "A Meta-Analysis of Research on Hindsight Bias." *Basic and Applied Social Psychology,* Vol. 26, 103-117.

[20] Hastie, R., D. Schkade and J. Payne. (1999), "Juror judgements in civil cases: Hindsight effects on judgements of liability for punitive damages." *Law and Human Behavior,* Vol. 23, 597-614.

[21] Harris, M. and B. Holmström. (1982), "A Theory of Wage Dynamics." *Review of Economic Studies,* Vol. 49, 315-333.

[22] Heath, C., and N. Staudenmayer. (2000), Coordination Neglect: How Lay Theories of Organizing Complicate Coordination in Organizations *Research in Organizational Behaviour,* Vol. 22, 155–193.

[23] Heath, C., R. Larrick and J. Klayman. (1998), "Cognitive Repairs: How Organizational Practicies can Compensate for Individual Shortcomings." *Research in Organizational Behavior*, Vol. 20, 1-37.

[24] Holmström, B. (1982), "Managerial Incentive Problems–A Dynamic Perspective." published in *Review of Economic Studies*, 1999. Vol. 66, 169-82.

[25] Jenter, D. and F. Kanaan. (2011), "CEO Turnover and Relative Performance Evaluation" forthcoming *Journal of Finance*

[26] Keysar, B. and A. Henly. (2002), "Speakers' Overestimation of Their Effectiveness." *Psychological Science*, Vol. 13, 207–212.

[27] Kessler, D. and M. McClellan. (1996), "Do Doctors Practice Defense Medicine?" *Quarterly Journal of Economics,* Vol. 111, 353-390.

[28] Kessler, D. and M. McClellan. (2000), "*How Liability Reform Affects Medical Productivity,*" *Journal of Health Economics*, Vol. 21, 931–955.

[29] Koessler, F. and P. Jehiel. (2008), "Revisiting games of incomplete information with analogy-based expectations." *Games and Economic Behavior*, Vol. 62, 533–557.

[30] Kruger, J., E. Nicholas, J. Parker and Z.Ng. (2005), "Egocentrism over E-mail: Can People Communicate as well as They Think?" *Journal of Personality and Social Psychology,* Vol. 89, 925-936.

[31] Loewenstein, G., T. O'Donoghue and M. Rabin. (2003), "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics,* Vol. 4, 1209-1248.

[32] Loewenstein, G., D. Moore and R. Weber. (2006), "Misperceiving the Value of Information in Predicting the Performance of Others." *Experimental Economics,* Vol. 9, 281-295.

[33] Madarász, K. (2009), "Information Projection: Model and Applications." mimeo, London School of Economics, London, UK.

[34] Madarász, K. (2011), "Projection Equilibrium in Bayesian Games." mimeo, London School of Economics, London, UK.

[35] Mandel, G. (2006), "Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational." *Ohio State Law Journal,* Vol. 67, 1391-1463.

[36] Mangelsdorff, L. and M. Weber. (1998), "Hindsight Bias im Prinzial-Agent-Konetxt: Die Aktennotiz als Antwort?" im H. Glaser, E. Schröder, A. v. Werder, eds. *Organisation im Wandel der Märkte*. Wiesbaden, Germany.

[37] Milgrom, P. (1981), "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics,* Vol. 12, 380-391.

[38] Newton, E. (1990), "Overconfidence in the Communication of Intent: Heard and Unheard melodies." Ph.D. Thesis. Stanford University, Stanford, CA.

[39] Pohl R., M. Eisenhauer, O. Hardt and J. Liebig. (2003), "SARA: A cognitive process model to simulate the anchoring effect and hindsight bias." *Memory,* Vol. 11, 337-356.

[40] Piaget, J. and B. Inhelder. (1967), *The Child's Conception of Space*. New York, Norton.

[41] Prendergast, C. and R. Topel. (1996), "Favoritism in Organizations." *Journal of Political Economy*, Vol. 104, 958-978.

[42] Rachlinski, J. (1998), "A Positive Psychological Theory of Judging in Hindsight." *The University of Chicago Law Review*, Vol. 65, 571-625.

[43] Radner, R. and J. Stiglitz. (1984), "A Nonconcavity in the Value of Information." in M. Boyer and R. Khilstrom eds. *Bayesian Models of Economic Theory*. Elsevier: Amsterdam

[44] Studdert, D., M. Mello, W. Sage, C. DesRoches, J. Peugh, K. Zapert and T. Brennan. (2005), "Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment." *Journal of the American Medical Association,* Vol. 293, 2609-2617.

[45] Wolfers, J. (2007), "Are Voters Rational? Evidence from Gubernatorial Elections." Working Paper Wharton School, UPenn, PA.