Information Projection: Model and Applications[1]

Kristóf Madarász

STICERD, LSE

First Version: December 2007. Current Version: November 2009.

[1] I thank George Akerlof, Jean-Pierre Benoit, Colin Camerer, Gary Charness, Marina Halac, Daniel Kahneman, Botond Koszegi, Ulrike Malmendier, Andrea Prat, Marit Rehavi, Matthew Rabin, Joel Sobel, Adam Szeidl and seminar participants at Stanford, UC Berkeley, LSE, UC San Diego, London Business School, Oxford, Yahoo! Research, Cambridge, IESE Barcelona, University College London, IZA Bonn, UC Santa Barbara and the Central European University for helpful comments. All errors are mine.

Abstract

People greatly exaggerate the extent to which their information is shared with others. I present a general model of such *information projection*, and apply it to a variety of settings. When assessing an expert's competence using ex-post information, jurors overweigh how much they learn from failed and underweigh how much they learn from successful predictions, and they *underestimate* the competence of experts on average. In turn, even risk-neutral experts are too reluctant to base predictions on ex-ante information that complements, and too eager to base predictions on ex-ante information that substitutes, for what jurors independently learn ex-post. Greater scrutiny decreases the reputation of physicians and the productivity of the medicine practiced at the same time. An expert's fear of information projection reduces her incentives to exercise care, and negligence schemes that tie her compensation too closely to effort often backfire. Optimal monitoring is coarser and career incentives are typically weaker than under fully Bayesian assumptions. Communication protocols that nudge experts to talk, but restrict the use of messages that complement the speaker's expertise, reduce favoritism and strictly improve welfare.

Keywords: Biased Beliefs, Optimal Monitoring, Defensive Medicine, Communication Protocols.

1 Introduction

Evidence from economics and psychology shows that people systematically *exaggerate* the extent to which their information is shared with others, and too often behave as if others guessed their private information correctly. Learning about symptoms a patient developed recently, jurors exaggerate the likelihood with which a careful physician should have detected cancer earlier. Similarly, though they have every incentive to communicate clearly, producers of electronic devices supply user-manuals that are too vague for their typical customer to understand. When advising graduate students, John Cochrane (2005) makes the following observation: "the most important thing in writing is to keep track of what your reader knows and doesn't know. Most Ph.D. students assume far too much. No, we do not have the details of every paper ever written in our heads."[1]

This paper provides robust evidence for and introduces a model of such *information projection*: people are aware of informational differences, but when imagining what others might know, people exaggerate the extent to which their information is available to others. People fail to appreciate the extent to which it is their private information that shapes their own beliefs, and too often assume that others already know what they do. Information projection will cause observers to blame experts too much in hindsight, and fear of projection will induce physicians or managers to adopt defensive practices when producing clinical tests or gathering investment data ex-ante. Since the presence of private information plays a central role in economic analysis, Akerlof (1970) and Spence (1973), the fact that people systematically mispredict informational differences is likely to have important consequences.

Section 2 reviews extensive laboratory and field evidence to support the existence of information projection. A variety of robust judgemental biases demonstrate information projection in a great number of domains. In the economic literature, Colin Camerer, George Loewenstein and Martin Weber (1989) were the first to provide directly relevant evidence. Significant information projection is present in the judgement of legal and medical professionals solving tasks they are thoroughly familiar with, which suggests that the nature of various social institutions might be affected.

Section 3, presents the model. I consider an environment where people obtain signals about an underlying state. If Alice suffers from information projection, she exaggerates the probability that the content of her signals is also available to Bob. Alice then mispredicts the distribution of information in the population. A key feature of this misprediction is that even if she is not strictly better informed, Alice always exaggerates *the value of* Bob's information. In particular, Alice overestimates how well Bob can do in expected utility terms and does so in proportion to how

[1] faculty.chicagobooth.edu/john.cochrane/research/.../phd_paper_writing.pdf

valuable her private information would have been to him. This identifying property of the model captures a variety of social mispredictions and distinguishes information projection from earlier anchoring based explanation offered for a subset of these mispredictions. Finally, it offers a unified framework to study their consequences both theoretically and empirically.

To illustrate the results, consider a simple medical example. A radiologist diagnoses a patient based on an ambiguous X-ray. Radiologists differ in their ability to interpret the X-ray, the best ones hardly ever miss a tumor when its visible, bad ones often do. Some time after the diagnosis is made, an evaluator is asked to review the radiologist's diagnosis. At this point, the evaluator has access to significantly more information than the radiologist did at the time of the original diagnosis. Interim medical outcomes are realized, new X-rays might have been ordered. A biased examiner projects this information, and believes that it was available for the radiologist at the time his original diagnosis was prepared. A very small tumor is typically extremely difficult to spot on an original X-ray, but once the location of a major tumor later is known, all radiologist have a much better chance of finding the small tumor on the original X-ray. The evaluator is now too surprised observing a failure and takes success too much to be the norm. As a consequence, she over-weighs how much failure reveals about competence and under-weighs how much success does and in turn underestimates the radiologist's competence *on average*.

Section 4 applies the above logic to skill assessment. In contrast to the Bayesian case, performance measures that are finer in ex-post information lead to lower assessments. When in the hope of obtaining more precise assessments, corporate boards decide to increase the scrutiny of CEOs, or government decides to investigate health professionals in more detail, they will consistently be disappointed in these experts. If lower average assessments imply a greater turnover of CEOs or a greater willingness to regulate health professionals, periods characterized by finer monitoring technologies are likely to tempt excess turnover and excess regulation.

Importantly, average underestimation does not imply undue skepticism after a success or a failure. To identify projection bias in conditional assessments, one needs to consider the *kind* of information projected. If the interim medical outcomes *complement* the skillful reading of the X-ray, the correct interpretation of the X-ray becomes more important in hindsight, and performance differences that arise due to luck are perceived as differences due to talent. In contrast, if interim medical outcomes *substitute* for the skillful reading of the radiographs, the evaluator believes that all types of radiologists had fairly equal chances of offering a successful diagnosis, and a bad outcome is interpreted as very bad luck. Hence differences in performance are attributed too much to differences in luck rather than talent. Thus, depending on whether more talented or less talented CEOs could have benefited more from knowing market outcomes in advance, CEO compensations will be too high after a success or not sufficiently low after a failure.

The most severe consequence of information projection happens, however, when experts *anticipate* biased evaluations. While economists have paid relatively little attention to projection bias, evidence from law and medicine suggests that professionals often recognize evaluators' tendency to project information and alter their behavior in response. Indeed in medicine, the fear from the *'retrospectroscope'* refers exactly to such recognition. To understand experts' responses, I turn to the case where experts have some discretion over what ex-ante information to base their predictions on in Section 5.

Even if he is risk-neutral, the radiologist distorts the production of information ex-ante in two opposite ways. If an ex-ante test reveals information that *substitutes* for the information that evaluators independently learn ex-post, then a physician has a preference to produce this test even in medically unwarranted situations. Overly costly or unnecessarily painful tests will be ordered to obtain information ex-ante that evaluators will independently learn ex-post. More surprisingly, the physician has a preference not to order medically efficient tests if these complement what the evaluators independently learn ex-post. A noisy mammogram might be the best, but still noisy, way to predict breast cancer at an early stage. Looking at the mammogram ex-post might allow the evaluator to determine with certainty whether a tumor was already developing ex-ante. Such ex-post insight, however, is often impossible in the absence of this ex-ante mammogram. Ironically, the radiologist can avoid underestimation and enjoy a higher expected reputation if he stays away from ordering such tests.

Production distortions occur in contexts where there are no agency conflicts under Bayesian assumptions. Under such conflicts, stronger production incentives might correct for distortions in the volume of the ex-ante information, when experts do not internalize production cost. Similarly, better risk-sharing could improve the skill-intensity of the ex-ante information, when radiologists are risk-averse. These changes, however, will not eliminate inefficiency here. Performance measures that are finer in ex-post information will shift the *composition* of the ex-ante produced from complement to substitute signals. To improve efficiency, either limited access to ex-post information or differential incentives for the production of substitute and complement information are necessary.

The results parallel stylized evidence from medicine. It has been repeatedly argued that it is jurors' fear of the ex-post exaggerations of the ex-ante accuracy of tests which is the prime motivation for 'defensive medicine'; medical practices adopted to minimize false liability, rather than maximize cost-effective health care, Leonard Berlin and Ronald Hendrix (1998). A vast majority of physicians engage both in the over- and in the under-production of skill-intensive medical tests to avoid bad reputation, David Studdert et al. (2005). Consistent with the model, weaker career concerns change the composition of diagnostic procedures in way that lowers medical costs but not the over-all quality of care, Daniel Kessler and Mark McClellan (1996, 2000).

In Section 6, I turn to the case where performance depends on effort rather skill to study how the results apply to contexts with explicit agency conflicts. If experts do not fully internalize the cost of careful information processing, an ex-ante commitment to evaluate performance ex-post is necessary for providing second-best incentives. Classic results on monitoring, Bengt Holmström (1979) and Sanford Grossman and Oliver Hart (1983), imply that finer performance measures will decrease the cost of agency conflicts and improve efficiency. An evaluator who suffers from information projection will commit two types of mistakes that are not expected under Bayesian assumptions. Because she judges ex-ante diagnosis too much on the basis of ex-post information, she punishes correct decisions too often and rewards bad decisions too often. Both types of judgemental errors mean that the expert's return on exerting care is diminished. Physicians who fear information projection thus *lower* the effort they exert to process information carefully relative to the Bayesian case.

These mistakes in monitoring imply that a transition from a strict liability rule to a negligence contract that should result in Pareto superior care under Bayesian assumptions will typically result in Pareto inferior outcomes under biased evaluations. A principal who designs incentives and understands the presence of information projection however, might be able to use biased reports in a way that restores the second best. I show that generally this is not the case. In effect, the reports of a biased evaluator contains too much *noise,* and distort the trade-off between rent extraction and incentive provision. As a result, even if the hospital is perfectly aware of evaluators' tendency to project information, it *cannot* eliminate the impact of this mistake. Optimal incentive schemes involve less monitoring, and when they opt for close monitoring, they induce lower care to save on incentives that are appropriate in the Bayesian case, but are too strong in the biased one.

In Section 7, I turn to the issue of communication. Information projection causes experts to send messages that are too ambiguous for lay audiences to understand. To the extent that they understand these messages, information projection causes listeners to over-conform to the advice they receive. The model predicts that protocols that mandate communication but also restrict the use of messages that complement the speaker's expertise dominate free-form communication and reduce miscoordination and mistaken favoritism in organizations.

Section 8 concludes the paper. I discuss applications to social conflict and self-control problems, an extension of the model to cover *ignorance projection*, and important limitations of my approach.

2 Evidence and Related Literature

This section presents evidence from various domains. Laboratory evidence on the curse of knowledge, Elizabeth Newton (1990), hindsight bias, Baruch Fischhoff (1975), illusion of transparency Thomas Gilovich et al. (1998, 2000) demonstrate the robust information projection in everyday social judgements. Field evidence with medical, business and legal professionals who solve tasks they are thoroughly familiar with provides further support. In fact, all the robust evidence in the interpersonal domain I am aware of is consistent with information projection, and while individual studies are often subject to alternative interpretations, the *sum-total* of the evidence provides a compelling case for presence of this mistake.

Informational differences play a key role in communication. In a striking study, Newton (1990) randomly assigned subjects at Stanford to be tappers or listeners and presented them with 25 well-known songs. Tappers had to privately pick one of the songs and tap out its rhythm. Listeners then guessed the song based on the rhythm. Out of 120 songs tapped, only three (2.5%) were correctly identified. When tappers were asked to predict these odds after picking their songs however, the mean prediction was around 50%. Thus in a context they should be familiar with, people projected their knowledge of the songs and overestimated the true faction of correct guesses twenty-fold. Similarly strong overestimation has been documented in a great number of recent studies, Chip Heath and Nancy Staudenmayer (2000), Boaz Keysar and Ann Henly (2002), Justin Kruger et al. (2005). A Heath and Heath (2007) write: "When a CEO discusses "unlocking shareholder value," there is a tune playing in her head that the employees can't hear."

While information projection is a very plausible explanation of the above result, people may simply overestimate their ability to communicate. Yet the same effect is demonstrated in various other domains without communication. In a fully *incentivized* study, Loewenstein, Don Moore, and Roberto Weber (2006) presented business students at CMU with verbal and visual tasks. In one set of tasks, subjects first saw two pictures that differed in one important detail. Subjects were divided into three groups: uninformed who received no further information, informed who were told the difference, and subjects who could choose to learn the difference for a smaller fee. In all treatments subjects had to guess the fraction of people in the *uninformed condition* who would correctly identify the difference and 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 the uninformed, and a significant share of subjects paid to learn the difference which then pushed their estimates further away from the truth (55% versus 30%). Since more

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, Don Moore, and Roberto Weber (2006)

information should help at least on average, informed estimates should have been closer to the truth. In this experiment, people not only projected their superior information, but paid for information that biased their judgements and systematically lowered their earnings.[2] In a similar experiment, raters who knew the solution to a word puzzle estimated whether solvers would figure out the solution, Emily Pronin, Carolyn Puccio and Lee Ross (2002). Here raters significantly overestimated the likelihood of success (83% as opposed to 21%), and attributed the difference to raters' low skill at puzzles.[3]

In the context of financial markets, Camerer, Loewenstein, and Weber (1989) provide further careful evidence with MBAs from Wharton and Chicago who trade assets via a double-oral auction. In one group traders learned the past performance of the traded companies and returns on trading where determined by the actual earnings of these companies. In the second group, traders also received the actual earnings, but here returns were determined by the market price established by the first group. In the second group, the market price was biased by 30% towards actual earnings. Individual judgements were biased by 60% . Though both largely significant, traders with a smaller bias traded more aggressively, reducing the bias in the market by acting *as if*, they anticipated the bias of others.

The most extensively documented form of information projection is the *hindsight bias*: after people are privately presented with a piece of novel information and asked to estimate what others should believe without it, they underestimate the importance of this information in shaping their beliefs, and report estimates that are systematically biased towards their private information. The first systematic demonstration of this fact is due to Fischhoff (1975). Fischhoff (1975) showed that reporting the outcome of an uncertain historical event changes the perceived ex-ante likelihood of the reported outcome occurring. A large literature following Fischhoff's study has shown that people report systematically biased estimates exactly in this direction of the hindsight effect. Rebecca

[2] To control for curiosity, LMW told subjects that they will learn the solution to the visual and verbal puzzles at the end of the experiment.

[3] Heath and Staudenmayer (2000) who replicated Newton's experiment found that over 40% of tappers attributed the surprisingly low success rates of listeners to listeners significant lack of effort when listening to their tapping rather than the difficulty of the task.

Guilbault et al. (2004) conduct a recent meta-analysis using 95 studies (83 published and 12 unpublished) including 252 independent effect sizes and document a very significant average hindsight effect under both objective and subjective uncertainty.[4] Hindsight bias is robust to a variety of debiasing techniques, e.g., Lawrence Sana, Norbert Schwartz and Shevaun Stocker (2002), and more ex-post information typically leads to greater absolute mispredictions, e.g., Pamela Hinds (1999).

Bayesian reasoning suggests that ex-post information can be valuable when interpreting ex-ante decisions. Hindsight bias suggests that people confuse the ex-post perspective with the ex-ante one. In a clean study Jonathan Baron and John Hersey (1988) tried to separate the second effect from the first, to demonstrate the presence of a hindsight effect in performance evaluation. Students at UPenn were asked to rate the quality of thinking that went into ex-ante decisions. Raters saw ex-ante decisions between a sure prize and a risky monetary gamble, e.g., \$100 for sure or a 50/50 chance of gaining \$220 or \$0, but also learned the ex-post realization of the gambles. They were told that these realizations were determined by the spin of a balanced roulette wheel after choices were made. Comparing 160 pairs of *ex-ante identical* choices, a higher ex-post earning was rated as a more correct ex-ante choice in 60% of cases, as an equally correct choice in 28%, and as a less correct choice in 12%. The same results were true when only forgone, not actual earnings were different.

The above findings are observed in economically much more important choices of professionals acting in their own fields. In medicine, Arkes, Saville, Wortmann and Harkness (1981) divided 75 practising physicians into five groups, gave them the same medical case history, and asked them to assign a probability estimate to each of four possible diagnoses being correct. One group received no additional information, but four "hindsight" groups were told an actual outcome. Nevertheless, all groups were asked to make their assessments based purely on the case history and independently of the diagnosis that turned out to be correct. The hindsight groups that were told that the least likely diagnoses were correct, assigned far greater probability estimates to these diagnoses than did the other groups. Caplan et al. (1991) present very similar evidence using 112 practising anesthesiologists and show that the difference in ruling ex-ante negligence can be as great as 51% between hindsight and foresight groups. Arkes et al. (1988) demonstrates the effect in a pool of 194 neuropsychologists, and Berlin (2000, 2004) in radiology.

In law, Anderson et al. (1997) demonstrate the mistake with practicing judges deciding on cases of auditors' liability in predicting the financial problems of their audit clients. Jeffrey Rachlinski (1998) and Erin Harley (2007) provide excellent surveys, Rachlinski (1998) and Christine Jolls,

[4] Other less recent meta-analyses include Christensen-Szalanski and Willham (1991) or Hawkins and Hastie (1990).

Cass Sunstein and Richard Thaler (2000) discuss the institutional responses in the legal code of the United States.

A set of other psychological mispredictions indicates that people project various forms of private information about themselves. Gilovich, Victoria Medvec, and Kenneth Savitsky (1998) provide clean evidence that people suffer from what they call the *illusion of transparency* or *spotlight effect* as they greatly overestimate the probability that their emotional and mental states are detected by others, or that their lies, once made, would be discovered.[5] Similarly, Gilovich, Savitsky and Medvec (2000) show that the average person overestimates the probability that others notice and the probability that others recall her actions and appearances.[6]

2.1 Related Literature

The closest to my paper is CLW (1989). They introduce an incomplete model of anchoring to illustrate their classic experimental results. They assume that better informed people are cursed because they are anchored too much to the *mean* of their own belief when imagining what a lesser informed person might know. Under specific distributional assumptions, Bruno Biais, and Weber (2008) follow the anchoring approach of CLW to offer a model of *intrapersonal* hindsight bias. They assume that people correctly remember the variance of their past beliefs, but misremember the mean, and specify conditions under which this leads to under-reacting to financial news. They test this prediction using psychometric and investment data on investment bankers from London and Frankfurt. In Section 3, I briefly compare anchoring assumptions with the mechanism of information projection. I show that an anchoring-based account is generally inconsistent with informational projection: a person who is anchored to his own beliefs would typically violate information projection and vice versa.

In the context of predicting future changes in one's own taste, the phenomenon of projection has been studied by Loewenstein, Ted O'Donoghue, and Matthew Rabin (2003). In contrast to the projection of taste, the projection of information is most relevant in the interpersonal domain and hence it is primarily a social bias. Finally, the paper complements a recent literature on limited strategic reasoning in Bayesian games with people who though predict information differences correctly, fail to appreciate the extent to which others condition their choices on their own private

[5] Van Boven, Gilovich, and Medvec (2003), study illusion of transparency in bargaining but these results are harder to interpret.

[6] The origins of information projection might have to do with people's limited ability of '*perspective-taking*'. Indeed Piaget's highly original theory of cognitive development is linked to a child's increased ability of perspective-taking over time, Inhelder and Piaget (1958). False-belief experiments were designed to demonstrate this, by showing radical information projection in children, Baron-Cohen (1985). Excellent recent psychological research shows, however, that the difference between adults' and children's' propensity to project information is not radically different. Susan Birch and Paul Bloom Paul (2003, 2007) show that adults also commit the false belief mistake in very simple tasks. Daniel Bernstein et al. (2004) show that interpersonal hindsight bias is not significantly diminished from children to adults.

information, Erik Eyster and Rabin (2005), Vincent Crawford and Nagori Iribberi (2007), Ignacio Esponda (2008), Eyster and Rabin (2009).[7]

3 Model

Consider an environment where people privately observe signals about the payoff-relevant state such as the fundamental value of a company or the medical conditions of a patient, $\omega \in \Omega$. There is a finite set of signals N and people M. The information of person $k \in M$ is the set of signals $I_k \subset N$ whose realizations she knows. With a slight abuse of notation, I denote both the set of signals and its power set by N. A signal is a function from the 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 finite set Ω.

To characterize the distribution of information, let $p_k^j \in [0,1]$ 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 p. This vector p describes the over-all distribution of information in this environment as it characterizes a complete probability distribution over the possible information sets in the population. Note that each sub-vector $p_k = \{p_k^j\}_{j=1}^N$ is a distribution over the subsets of N assigning a probability to the events that person k faces information I_k for all $I_k \subset N$. In turn, the informational environment can be summarized by the tuple $\Gamma = \{\Omega, \sigma, \{s_j\}_{j=1}^N, p\}$. To conclude the setup, let person k's finite action set be Y_k, and her von-Neumann-Morgenstern utility function $u_k(y,\omega) : Y_k \times \Omega \to \mathbb{R}$, where $y_k \in Y_k$.

3.1 Definition

As long as people have rational expectations, the vector p describes people's perception of how information is distributed. Information projection introduces a *bias* in the perception of this vector p. A person who projects information, exaggerates the probability that a signal that is in her information set is also in the information set of others. I introduce a parameter $\rho \in [0,1]$ to express the degree of such information projection.

Definition 1 *Person k with information set I_k exhibits information project of degree ρ if her perception of the distribution of person i's, $i \neq k$, information p_i is given by the vector p_i^ρ where*

$$p_i^{j,\rho} = (1-\rho)p_i^j + \rho \quad \textit{if} \quad s_j \in I_k \quad \textit{and} \quad p_i^{j,\rho} = p_i^j \quad \textit{if } s_j \notin I_k \quad \textit{for any } j \in N \tag{1}$$

[7]Several other papers, with no explicitly developed model or results, argued about the potential importance of various mistakes one can classify as information projection, e.g., Christine Jolls, Cass Sunstein and Richard Thaler (2000), Viscusi, and Zeckhauser (2005), Camerer, and Ulrike Malmendier (2007).

In the case of full information projection, $\rho = 1$, a biased person believes that all her information is available to 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 information projection case. Finally, when $\rho = 0$, she has correct Bayesian expectations.[8]

The degree of projection is uniform in the above definition, an assumption made for notational simplicity only. Generally, the degree of projection can be represented by a vector, rather than a scalar where different signals can be projected to varying degrees. Formally, if ρ_k^j is the degree to which person k projects signal j, then $\rho_k = \{\rho_k^j\}_{j=1}^N$ is person k's generalized degree of projection. Although in some applications predictions might depend on whether the degree of projection is homogenous or heterogenous this is not true for the results here. All results of the paper extend to heterogenous projection and whenever I refer to an increase in the bias, I mean an increase in *any* component of this vector. The claim of the model is only that $\rho_k^j \geq 0$. The above definition is formulated without explicit reference to time. If $i' \in M$ is the past or future self of person $i \in M$, the definition claims that person k projects her *current* information onto this past or future self of person i.[9]

3.2 Example

To illustrate the model, consider the following example where a host invites a guest over for dinner. The host can either prepare fish or meat. If he is kind, the host wants to prepare the guest's favorite. If he is unkind, he prepares what he likes the most. Suppose that ex-ante the host is equally likely to be one of four possible types θ: $\{kind, mean\} \times \{similar, different\}$. The guest initially believes that the host is equally likely to be kind or unkind, and have the same or the opposite preference as she does.

Suppose that while the guest knows her taste, the host is uncertain about this, and receives only noisy information only. In the example below, this information conveys the guest's correct preference ranking 2/3 of the time and the incorrect one 1/3 of the time. A Bayesian guest fully appreciates the informational asymmetry, a biased guest projects her information acting as if her

[8]The model can be extended by allowing the parameters p to depend on the state ω, $p(\omega)$. In this formulation, after observing a set of signals, a Bayesian person forms a 'posterior estimate $p_i(\omega)$ denoted by $p_i^e(\omega)$. The definition then can be applied to this vector $p_i^e(\omega)$ 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.

[9]It is important to note that while the above definition adopts a simple linear form, all that matters for the theory of information projection and for the qualitative results of this paper that $p_i^{j,\rho}$ is continuously increasing in ρ. Also, the above definition is appropriate for simpler Bayesian problems, but when considering more complex Bayesian games or limit results, to account for the appropriate measurability conditions, it might be more appropriate to adopt the definition where $p_i^{j,\rho} = (p_k + \rho)/(1 + \rho p_k)$ if $s_j \in I_k$ and p_j^i otherwise. The only key difference here, is that if $p_j^i = 0$, then $p_i^{j,\rho} = 0$ for all ρ.

taste was public information. The following table summarizes the respective inferences:

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

Two types of inferential mistakes are present in this situation that play an important role in the paper: misattributions and underestimation. Note first, that a biased guest *overinfers* kindness if she is served the meal she likes, and she overinfers hostility if she is served the meal she dislikes. These two effects do not cancel out. On average, the guest believes that her host is mean, $\frac{7}{18} < \frac{1}{2}$.

Similar calculations show that the guest under-infers taste differences from the host's meal choice. Where a Bayesian guest would learn about the taste of her host, a biased one learns about the host's intentions towards her. More importantly, however, on average, a biased guest also underestimates the probability that her host has a similar taste to hers, $\frac{7}{18} < \frac{1}{2}$. To understand these result more carefully, let me now turn to the identifying property of the model.

3.3 Projection and the Value of Information

Below, I first show that a person who suffers from information projection exaggerates the *value of information* others have and does so in proportion to how valuable her information would have been to them. I then identify a class of monotone inference problems where a biased observer, learning about an actor, comes to *underestimate* the productive qualities of the actor *on average*. I proceed in a more abstract manner. A reader not interested in the more general formulation, should skip the details below, and will find a variant of the results presented below in Section 4 along with the economic intuition.

Estimating Expected Utility. Consider the expected utility maximization problem of person i. To economize on notation, in what follows I drop the index i. Let $u^*_I \in \mathbb{R}$ denote the value of this program when person i has information $I \subset N$. Formally,

$$u^*_I = \max_y E[u(y, \omega) \mid I] \tag{2}$$

Note that from an ex-ante perspective, u^*_I is a random variable in environment Γ whose distribution is determined by $\underline{p}$. Accordingly, let $f^\rho(u^*) \in \Delta\mathbb{R}$ denote the pdf which stands for the another ρ-biased person's belief about the distribution of u^*_I. Such beliefs will typically depend the projector's information I_k. The following result shows a general property of these beliefs independent of the details of the information set I_k.

Proposition 1 *Given vN-M preferences in environment* Γ*,* $f^{\rho}(u^*)$ *first-order stochastically dominates* $f^{\rho'}(u^*)$ *if and only if* $\rho < \rho'$ *for all* I_k*.*

A biased person exaggerates how informed others are and as a result overestimates how well they can do in expected utility terms. The proof of this result is based on Blackwell's classic theorem on the comparison of information sets, Blackwell (1953). While Blackwell offers only a partial ordering of information sets, the proof of the above proposition shows that the misperception induced by information projection can be ordered by this criterion. A corollary of this result is that the more a projector knows, the higher is his estimate of how well others can do in expected utility terms.

Corollary 1 *An increase in the Blackwell informativeness of* I_k *, leads to an increase in* $f^{\rho}(u^*)$ *in the sense of first-order stochastic dominance if and only if* $\rho > 0$*.*

Inference. Let me now apply the above result to a class of economic inference problems. An observer is learning about the hidden type θ of an actor by observing x. The outcome is an element of a finite ordered set X and can be understood as some noisy signal of the actor's actions or intentions. The actor's type is also an element of an ordered set Θ, and will be interpreted as his competence, fairness, loyalty, etc. The key assumption is that the outcome x depends positively both on the actor's type, θ, and the value of his information, u^*. Formally, let there be a family of real valued conditional density functions, $\{f(x \mid \theta, u^*)\}$, that satisfies the monotone-likelihood-ratio property given priors $f(\theta)$ and $f(u^*)$.

Condition 1 (MLRP) *Let* $f(x \mid u^*) = \int_{\Theta} f(x \mid \theta, u^*)d\theta$ *and define* $f(x \mid \theta)$ *analogously. The family* $f(x \mid \theta, u^*)$ *satisfies the strict monotone likelihood ratio property in both* θ *and* u^* *if*

$$f(x \mid u^*) * f(x' \mid u^{*\prime}) - f(x \mid u^{*\prime}) * f(x' \mid u^*) > 0 \tag{3}$$

whenever $x > x'$ *and* $u^* > u^{*\prime}$*, and the analogous condition holds for* $\theta \in \Theta$*.*

Proposition 1 claims that a biased observer exaggerates the value of the actor's information. This means that on average she expects a higher outcome than what is actually realized. Because her expectations about the level of x are not met, she revises her beliefs about the quality of the decision-maker downwards. Formally, the beliefs of a biased observer's belief about u^* are given by $f^{\rho}(u^*)$. Let $f^{\rho}(\theta \mid x)$ denote her posterior on θ after observing x. The next result shows that as long as Condition 1 holds, the observer underestimates θ on average.

Proposition 2 *Suppose Condition 1 holds. For all priors* $f(\theta)$ *and* $f(u^*)$*,* $E_X[f^\rho(\theta \mid x)]$ *first-order stochastically dominates* $E_X[f^{\rho'}(\theta \mid x)]$ *whenever* $\rho < \rho'$ *for expectations taken with respect to the true distribution of signals.*

Above, I assumed that a higher outcome is *good news* about quality θ. Since the result depends only on the monotonicity assumption, it follows that when a higher outcome is *bad news* about θ – in the sense of Milgrom (1981) – a biased observer will overestimate θ on average, and her expected beliefs about θ will be too high. Finally, when x is neutral about θ, no misestimation is implied by information projection.

3.4 Discussion

The above results do not depend on details of the environment Γ. Specific restrictions might be imposed on the distribution of information, the partitioning of informational differences into distinct signals, or on the players' expected utility preferences. These enrich the set of predictions, but do not violate the above results.

Anchored Expectations or Projection Bias. A strictly binding restriction on Γ is the presence of a monotone ordering of information. In such a context, CLW (1989) explain the curse of knowledge by the following incomplete anchoring assumption: a better informed trader perceives the *mean* expectation of a less informed trader to be the convex combination of her mean expectation and this lesser informed trader's true *mean* expectation. The example below shows that the exaggeration of the *proximity of two means* is not a measure of informational closeness, and hence anchoring assumption violates the key property of information projection.

Example 1 *Let there be three people with strictly ordered information about the return on an asset.* Sam1 *is uniformed and has a uniform prior on* $[0, 1]$. Sam2 *receives valuable information and learns that the quality is either* 0 *or* $\frac{3}{4}$ *with equal probability.* Sam3 *learns that the true quality is* $\frac{3}{4}$*. The distance between* Sam1's *and* Sam3's *mean belief is* $\frac{2}{8}$ *while the distance between* Sam2's *and* Sam3's *mean belief is* $\frac{3}{8}$*.*

Anchored expectations imply neither the overestimation nor the underestimation of the value of another person's information. A person who projects information might exaggerate the distance in mean beliefs, as is often the case when complement information is projected. Thus the behavioral and welfare implications of the current model are very different from those that could followen from completing an anchoring based approach.

3.5 Substitute and Complement Information

Key implications of the model depend on the *kind* of information projected. A distinction between substitute and complement information is key. Two images of a bone fracture on an X-ray and on an MRI are substitutes, if any one of the two is broadly sufficient to establish the fracture. Knowledge of the location and the knowledge of the type of a tumor are complement signals however, if only the combination of these two allows one to identify the best treatment.

Holding constant an objective function, two signals are substitutes if knowing both is less valuable than the sum of knowing each separately. They are complements if the opposite holds. If again $u^*(s)$ is the value of a person's utility maximization problem given information s, signals are substitutes if u^* is *submodular* in them, and complements if u^* is *supermodular* in them.[10]

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

To conclude the setup of this model, I introduce the possibility that some people might correctly anticipate the bias of others. As mentioned in Section 2, evidence suggests that in important contexts, people do anticipate the bias and respond to them. Since I build on this in the applications, I define such anticipation formally. Let the probability density function $\psi_i(\rho_k)$ on $[0, 1]$ describe the beliefs of person i concerning the extent to which 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 Inference

Let's now turn to the main application of the paper. Consider an environment where a *supervisor* evaluates an *agent* whose task is to process and act upon information. Performance evaluations of this sort are often considered to be one of the key rationales for the existence of firms and play an important role in organizations and labor markets. Performance evaluations aid promotion decisions, job design, the efficient *allocation* of human resources more generally, and are essential in providing incentives when parties have conflicting objectives. When evaluating the agent ex-post, the supervisor typically has access to novel information that was not available ex-ante. This creates room for information projection. Section 4 considers inference problems where the goal of performance evaluation is a process of pure inference designed to learn about the skill of agents. Section 5 introduces career incentives into this setup and addresses the case where the agent anticipates biased evaluations. Section 6 maintains the assumption that agents anticipate biased evaluations,

[10] On the presence of non-concavities in the value of information see e.g., Radner and Stiglitz (1984).

but extends the setup to the case where performance evaluation also serves a commitment role in the provision of incentives once explicit agency conflicts are introduced.

4.1 Setup

In the beginning, the radiologist receives an ambiguous *X-ray* about the medical condition of a patient. Let's denote this noisy ex-ante signal about the state ω by s_0. Based on his information, the radiologist makes a prediction/adopts a treatment, $y_a \in Y_a$. Given a technology matrix A, that depends both on the state and on the radiologist's action, an outcome x is realized which is interpreted either as a success x_S or as a failure x_F.

Technology. Let $\theta \in [0,1]$ denote the radiologist's competence where θ denotes the probability that the radiologist understands the X-ray. If he does understand it, his beliefs about the state remain unchanged. More generally, I assume that for any fixed information set I_a, a more competent type understands a superset of the signals a less competent type does in a probabilistic sense. Let A be an unrestricted non-negative matrix with dimension $|Y_a| \times |\Omega|$ in which each element determines the probability that a particular action-state pair will give rise to success.[11]

Agent's Choice The radiologist's goal is to maximizes the probability of success. Given his beliefs σ_1 about ω, which incorporates the content of s_0 with probability θ, the radiologist picks an action $y_a^* \in Y_a$ to maximize

$$\max u_a(y_a, \omega) = \max y' A \sigma_1 \tag{4}$$

where y' is a non-negative $|Y_a|$ dimensional indicator vector with exactly one positive element.

Supervisor's Information The supervisor knows A and observes x. Since x depends on ω even after controlling for s_0 and the agent's action y_a, generically, the realization of x provides novel information about ω.[12] A realized medical outcome typically carries novel information about the medical condition of the patient. In addition, however, the supervisor might also have access to a new X-rays that have been ordered. I denote the total novel information about ω by s_1. Given her information, the supervisor updates her prior on θ, which I denote by π_0, via Bayes rule.

Equilibrium For the analysis below, specific assumptions on the objective of the supervisor are not necessary. It suffices to assume that she is an expected-utility maximizer who is interested in the true value of θ. It also *suffices* to assume that for extrinsic or intrinsic reasons, the radiologist prefers a success to a failure. To further motivate such behavior, suppose that the agent has career concerns and thus maximizes an increasing function of his expected reputation. The following lemma shows

[11] Based on Proposition 2, I conjecture that the main results of this section remain valid when one allows for the fact that the probability of understanding an ex-ante signal s_0 depends on the state ω as well as long as the MLRP is satisfied.

[12] For a small, but important set of technologies – e.g., where A is a symmetric diagonal matrix – this information might not be strict.

that in the efficient Bayesian Nash Equilibrium of the above game, his strategy is again to take an action that maximizes the probability of success. Lemma 2 will show that this remains true in the biased case as well.

Lemma 1 *Let $u_a(y_a,\omega) = E_{y,\omega}[b(\pi_1)]$ where $b(\pi_1) : \Delta[0,1] \to \mathbb{R}$ and $b(\pi_1) \geq b(\pi_1')$ whenever π_1 first-order stochastically dominates π_1'. For any π_0 with full support, the efficient Bayesian Nash Equilibrium involves y_a^* as given by Eq. (4).* [13]

4.2 Skill Assessment

Given her perception of the agent's ex-ante information and behavior, she updates her prior on θ, which I denote by π_0, via Bayes rule. Information projection introduces a bias in this perception. A biased supervisor projects the novel information about ω revealed along with x and forms a posterior accordingly. Let π_1^{ρ} denote the posterior of a ρ-biased supervisor. The following result adopts Proposition 2 to the case of binary outcomes.

Proposition 3 *For all A, π_0, s_0,s_1 $E_{\omega}[\pi_1^{\rho}]$ first-order stochastically dominates $E_{\omega}[\pi_1^{\rho'}]$ whenever $\rho' \geq \rho$ where expectations are taken with respect to the true distribution of signals.*

Intuitively, by projecting information, the supervisor overestimates the probability that the radiologist's prediction should be a success. As a result, she is too surprised observing a failure and less surprised observing a success than she should be and puts too much weight on the information revealed by the former and too little revealed by the latter. Since the probability of a failure is decreasing in the radiologist's competence, the supervisor underestimates his type on average. Note that the above result holds for all priors with full support, and hence when the supervisor samples from a longer sequence of performance, underestimation holds a fortiori.[14]

A comparative static result for the above proposition claims that the more useful the ex-post information would have been ex-ante, the greater is this *informational gap*, the more skeptical the supervisor will be on average.

Corollary 2 *Let $g = u_a^*(s_0 + s_1) - u_a^*(s_0)$. Then $E_{\omega}[\pi_1^{\rho}]$ is decreasing in g in the sense of fosd if and only if $\rho > 0$.*

[13] Since there is no special effort cost associated with any action in Y_a, the shape of b does not play a role in the analysis. This is unlike the classic signal-jamming models of implicit incentives e.g., Holmstrom (1982), Scharftstein and Stein (1990), Dewatripont, Jewitt and Tirole (1999).

[14] Under auxiliary measurability assumptions it can be ensured that biased estimates $\pi_t^{\rho}(\theta)$ comprise a *super-martingale* for all $\rho > 0$ and the $\lim_{t\to\infty} \pi_t^{\rho}(\theta \mid \widehat{\theta})$ does exists for a.a. $\widehat{\theta} \in [0,1]$ and in addition, $\lim_{t\to\infty} \pi_t^{\rho}(\theta \mid \widehat{\theta}) < \lim_{t\to\infty} \pi_t^{0}(\theta \mid \widehat{\theta})$ for all $\widehat{\theta} > 0$.

More Scrutiny →Lower Assessments To obtain more valuable estimates of the agent's competence, the supervisor should adopt the finer partitions of the ex-post available information in the Bayesian case. Importantly, due to the martingale property Bayesian updating, increasing the fineness of the ex-post information should not change the expected level of assessments, i.e., $E[\pi_1] = \pi_0$ should hold for all $s_1 \subset N$. In the biased case, adopting finer performance measures will *lower the level* of the posterior. The more novel information the supervisor learns about the task, the lower is the weight she puts on the information revealed by a success and the greater is the weight on the information revealed by a failure. Hence, when the boards decide to increase the scrutiny of CEOs or a government decides to investigate the activity of social workers in greater detail, the reputation of these experts will suffer. And when lower average assessments imply a greater probability of CEO turnover or a greater willingness to regulate social workers in our setup, the model makes the prediction that the decision to switch to monitoring technologies that involve finer ex-post information on the task solved by experts will cause too much turnover or too much regulation.

4.3 Relative Performance Evaluation

Let me further illustrate the implications of Corollary 2. Consider relative performance evaluations, a common practice in organizations. The supervisor now evaluates two agents working on independent tasks. Let the temporal information gap be g_1 for the first task and g_2 for the second. As long as the tasks are homogenous, $g_1 = g_2$, information projection will not affect the ranking of these two agents. As soon as $g_1 > g_2$ however, information projection can *reverse* long-term rankings due to the misattribution of information differences. This gives rise to favoritism and relative overconfidence.

Favoritism Assume that the two agents are equally competent, but the first solves a task with perfect ex-ante information and the second solves one where uncertainty fully resolves only ex-post. Here a biased supervisor will rank the first agent higher on average than the second.[15] More generally, using the terminology introduced in Definition 2, a perfectly neutral but biased supervisor will favor the agent whose ex-ante information is a *closer substitute* of her information. In this manner, informational differences produce an endogenous link between the distribution of information in an organization and mistaken favoritism. Existing accounts of this phenomenon, e.g., Prendergast and Topel (1996), link favoritism to an exogenous preference supervisors exhibit for some, but not all workers. The current model links favoritism to informational differences and predicts that harmonizing information gaps can lessen the scope of inefficient favoritism.

[15] Durell (1999) provides related laboratory evidence. Employees were randomly assigned to easy or hard word puzzles. Employers observed the performance of the employees on these tasks and had to predict their future performances. Employers correctly predicted the future performance of those initially assigned to easy tasks, but *underestimated* it for people initially assigned to hard tasks.

Overconfidence To further illustrate the above comparative static result, suppose someone is trying to predict her own competence θ and her future performance on an unsolved task $\{A, s_0\}$ by sampling equally difficult problems from past tasks $\{\widetilde{A}, \widetilde{s_0}\}$. If this person also receives the solution manual for this past task $\widetilde{s_1}$, she will *overestimate* how well she would do on the exam on average. The model predicts that such overconfidence is mitigated by changing informational conditions; people who prepare for the exam without a solution manual will make less exaggerated predictions of their future performance on average.

4.4 Luck or Talent

Note that underestimation on average does *not* imply undue pessimism either after a success or a failure. *Conditional assessments* depend on the *kind* of information projected. If the projected information had been more useful for high types than for low types, the supervisor overinfers skill from performance. If the reverse is true, she underinfers skill from performance. Formally, let $\pi^{\rho}(x_s \mid \theta)$ express the conditional probability of success as perceived by a ρ-biased supervisor.

Proposition 4 *If* $\frac{\pi^{\rho}(x_S|\theta)}{\pi^{0}(x_S|\theta)}$ *is increasing in* θ*, then* $\pi_1^{\rho}(\theta \mid x_S)$ *fosd* $\pi_1^{0}(\theta \mid x_S)$ *for all* π_0*. If* $\frac{\pi^{\rho}(x_S|\theta)}{\pi^{0}(x_S|\theta)}$ *is decreasing in* θ*, then* $\pi_1^{\rho}(\theta \mid x_S)$ *fosd* $\pi_1^{0}(\theta \mid x_S)$ *for all* π_0*.*

To illustrate the intuition, consider the special case, where ex-post information is not skill-intensive: had it been available ex-ante, all types would have understood it equally well. We have to distinguish between two cases: (i) the projected information is a complement of the ex-ante skill-intensive signal, (ii) the projected information is a substitute of the ex-ante signal. If the ex-post information is a complement, the supervisor misattributes differences in luck to differences in skill and overinfers skill from performance. Suppose that ex-post market conditions s_1 would have helped only a CEO who understood the quality of a firm's assets s_0. Here the supervisor exaggerates how much skill differences matter for performance and hence even when s_0 alone was uninformative, supervisors will infer skill from performance leading to an *illusion of talent.* [16]

In the case of projected substitute information, the supervisor down plays the role of skill differences in performance differences. In hindsight, only differences in luck can explain differences in outcomes because the task appears equally obvious for all types. Here, if the force of under-inference is greater than the force of underestimation, the projection of substitute information leads to an *unwarranted excuse for failure.*

[16] Consistent with this result, Jenter and Kanaan (2006) uses a sample of 1,627 CEO turnovers from 1993 to 2001 and they show that even when they should, boards do not filter observable exogenous shocks from firm performance before updating their assessment of CEO quality. Moreover, they find that bad market outcomes increase CEO dismissal in a fashion not explained by rational learning models or strategic considerations.

Lemma 2 *Assume* s_1 *requires no skill to process. If* s_1 *and* s_0 *are complements given* A*, then* $\pi_1^\rho(\theta \mid x_S)$ *is increasing in* ρ *and* $\pi_1^\rho(\theta \mid x_F)$ *is decreasing in* ρ *in the sense of fosd for all* π_0*. If* s_1 *and* s_0 *are substitutes given* A *, then* $\pi_1^\rho(\theta \mid x_S)$ *is decreasing in* ρ *in the sense fosd for all* π_0*.*

5 The Supply of Information

The analysis above established that information projection has an adverse effect on an expert's expected reputation. In light of this, let's turn to the case where the radiologist ex-ante *anticipates* biased evaluations and responds to them strategically.[17] To understand how the radiologist might change his behavior in response, note first for all beliefs about the supervisor's bias, the radiologist optimal choice is still given by Eq. (4).

Lemma 3 *For all* $\psi(\rho)$*, the agent's best response is given by Eq. (4).*

Although the radiologist still takes an action to maximize the probability of a successful outcome, he has non-standard preferences over what information to base his predictions on ex-ante. To analyze this scenario, assume that the agent can decide whether to produce an additional radiograph, s_0', or not. The social value of producing s_0' is given by $\mathbf{a} \in \mathbb{R}$.[18] To separate the analysis from any direct agency conflict, assume that the agent fully internalizes a. The timing is as follows: the agent privately observes a, publicly decides to produce s_0' or not, then takes an action y_a and finally x is realized. I assume that the production of s_0' is *perfectly observable* by the supervisor and that the agent is risk neutral over assessments, $b(E[\pi_1^\rho]) = E[b(\pi_1^\rho)]$. I relax both of these assumptions in the discussion below. Formally, the agent's objective is now

$$u_a(y, a, \omega) = \chi a + b(E_{y,\omega}[\pi_1^\rho]) \tag{5}$$

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

To make the comparative statics more pronounced let m be the probability that the agent is assessed. If he is not assessed his reputation remains unchanged. In the Bayesian case, independent of m, the agent produces s_0' whenever it is socially optimal. In contrast, an agent who fears information projection *over-produces* tests that *substitute* for, and *under-produces* tests that are *complements* of the evaluator's ex-post information.

[17] There is strong evidence that people display a strong asymmetry when thinking about the fact that others might be biased versus assessing the fact that they themselves are biased, Emily Pronin, Thomas Gilovich and Lee Ross (2004).

[18] Here benefits include the additional knowledge gained about the patient, and production costs include the alternative use of medical resources, increased exposure to pain/radiation, the delay in treatment.

Proposition 5 *Suppose the supervisor observes the production of s'_0. For all A, 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, 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 $\rho > 0$.*

Producing an additional substitute information reduces the information gap between the ex-ante and the ex-post stages. If such information is included in the set of ex-ante radiographs the radiologist bases his predictions on, he can expect a higher ex-post reputation In this manner, an otherwise fully concerned physician will produce such tests in medically unnecessary situations to limit the scope of false liability. More surprisingly, the opposite under-production happens when the ex-ante test complements the ex-post information. In this fashion, a social worker will avoid making a costless but valuable phone call to a foster family to avoid underestimation. Assume that ex-post evaluators will inevitably learn whether an accident has happened to the child, but this will only be informative of child abuse once combined with the information that only an ex-ante call can reveal. Here, the social worker can prevent such ex-post insight by not making the phone call ex-ante. To limit underestimation, he will indeed decide to do so.

Increased Monitoring → Less efficient Tests + Lower Reputation. Consider now what happens as the frequency of monitoring m is increased: the composition of the ex-ante produced information shifts from more efficient complement signals to less efficient substitute signals. This will lead to a decrease in the quality of the medicine practiced. At the same time, an increase in m can also lead to a decrease in the reputation of physicians. This joint effect identifies the above proposition. Before linking these predictions to stylized evidence from medicine, let me briefly turn to robustness when introducing risk aversion over assessments.

Risk Aversion versus Fear of Information Projection. To simplify the analysis, I assumed risk-neutrality above. As noted by Holmström (1982), a risk averse agent with career-concern might want to limit the amount of inference the supervisor makes about θ to limit reputational risk.[19] Indeed, if by not producing s'_0 the agent could eliminate all inference about his skill, i.e. in effect set $m = 0$, even a risk-neutral agent would want to do that. More importantly, because a risk-averse agent prefers a lottery that dominates another in the sense of fosd, he might well over-produce substitute information, even if this allows for more inference about his skill. For this reason, risk-aversion will typically *amplify both* the over-production of substitute and the under-production of

[19] On this classic effect of risk-aversion see e.g., Hermalin (1993).

complement information. Inefficiency in production arises even when optimal risk-sharing is not an issue and the amount of inference about the agent's skill is held constant.

Production Incentives. Let's now consider the case where the agent does not fully internalize a. Assuming that there is prior distribution on a, $f(a)$, stronger career incentives are necessary if the agent does not fully internalize the benefits of production, $a(0,0) > 0$, and weaker if he does not fully internalize the costs $a(0,0) < 0$. Importantly, stronger or weaker production incentives do not eliminate the distortions caused by information projection. The agent here still enjoys a higher expected utility if he distorts production from complement to substitute signals. The problem remains, unless *different incentives* are in place for the production of substitute and complement information.

Reputational Loss and Defensive Medicine. Consistent with the results, Studdert et al. (2005) interviewed physicians in 6 high-risk medical fields in the state of Pennsylvania. Of the 824 physicians interviewed, 93% reported that they engage in defensive medicine, 43% reported using imaging technology in clinically unnecessary circumstances, 42% reported that they had taken steps to restrict their practice and avoid using various efficient technologies such as emergency surgeries, complex obstetrics and mammograms. In radiology, Berlin (2000, 2004) argues that medical lawsuits stem from the serious misconception of juror's that radiologist should be able to see what may be discernible only in hindsight. Furthermore, Berlin (2000) and Jackson, and Righi (2006) report that it is due to such ex-post exaggeration of the accuracy of mammograms that radiologists are reluctant to adopt what is believed to be the best way to detect breast cancer early. Berlin provides evidence that while the generally accepted error rate for radiologic detection of lung cancer is between 20% and 50%, evaluating previously "normal" chest radiographs of patients who subsequently developed lung carcinoma reveal that the carcinoma could be seen in retrospect in as many as 90% of cases.

Supportive of the above setup, Kessler and Mclellan (1996) argue that the main motivation for defensive medicine is not pecuniary loss rather the fear from the loss of reputation, and show that reducing liability pressure leads to reductions in medical expenditures between 5% and 9 % without substantial effects on mortality or medical complications.[20] Supportive of the mechanism of information projection, Kessler and McClellan (2000) find that the main effect of cost reduction is on *diagnostic* rather than on therapeutic practices. Although defensive practices are sometimes

[20]Importantly, it is not the pecuniary loss that is important in medical malpractice. The compensation paid and the costs of administering that compensation through the legal system account for less than 1 percent of medical expenditures. As Kessler and McClellan (1996) argue however, "the effects of the malpractice system on physician behavior, in contrast, may have much more substantial effects on health care costs and outcomes, even though virtually all physicians are fully insured against the financial costs of malpractice such as damages and legal defense expenses. Physicians may employ costly precautionary treatments in order to avoid nonfinancial penalties such as fear of reputational harm, decreased self-esteem from adverse publicity".

attributed to physician's fear of random judicial judgements, if the cause of false liability is not random judgement but information projection, then increased efficiency should operate through an observable change in the *composition* of the diagnostic practices as specified by Proposition 5.

Unobservable $\mathbf{s_0'}$. The analysis above assumed that the supervisor observes the production of s_0 and hence when making inferences takes this decision into account. To link information production decisions to distortions in conditional assessment as established in Proposition 4, consider now an *out-of-equilibrium* scenario where s_0' is produced secretly, i.e., the supervisor does not observe the production of s_0' and in fact believes that s_0' could not have been produced.

Corollary 3 *Suppose s_1 requires no skill to process. If $\rho = 0$, then $a(0, m) < 0$. If s_1 and s_0 are complements given A, then the agent's willingness to produce s_0' is increasing in $\psi(\rho)$. If s_1 and s_0 are substitutes given A, then the agent's willingness to produce s_0' is decreasing in $\psi(\rho)$.*

The assumption that the supervisor assigned a zero probability to the production of s_0' simplified the presentation, but the same directional effects are true when the supervisor has a fixed expectation about the probability of producing s_0'. These incentives, in contrast to those determined by Proposition 5, are eliminated however, when the supervisor's expectations are conditioned on the true production choice.

6 Incentives

So far, the agent's performance was determined by his skill alone. Needless to say, in many scenarios performance also depends on the care of the expert in processing information. If the expert fails to fully internalize the benefits of careful information processing, a problem of moral hazard arises. Whenever care is not observable, an ex-ante commitment to evaluate the expert ex-post is needed to provide incentives to achieve the second-best. As it is well understood since Holmstorm (1979) and Grossman and Hart (1982), the efficiency of such incentive provision will depend on the optimal use of performance information.

Biased evaluations impact the theory of the second-best. First, the fear of biased evaluations prompts an expert to exercise uniformly less care. Second, Bayesian superior monitoring or a shift to Bayesian optimal negligence rule often backfires. Finally, even an incentive designer who predicts information projection perfectly, cannot eliminate its impact; rather, he decides to monitor less often and tie rewards less closely to effort. The analysis below complements the previous section. In Section 5, I endogenized the information gap between the expert and the supervisor. Here I hold the absolute size of this gap constant, but specify the agent incentives to reduce it by exerting effort. The results show that even if differential incentives are offered for the production of substitute and complement signals, finer monitoring still distorts the effective production of information.

6.1 Setup

After receiving information, but before selection a recommendation y_a, the expert decides how much care, $e \in \mathbb{R}^+$, to exert. His care determines the probability $p(e)$ that he understands s_0. I assume that $p(e)$ is a concave technology and that the usual limit conditions apply. The expert and the principal are risk neutral with $u_a(w, e) = w - e$ and $u_p(r, w) = x - w$, where x is worth 1 if it is a success, and 0 otherwise, and w is the agent's wage. Finally, I assume that the agent is protected by *limited liability*, $w \geq 0$ must hold, hence a classic trade-off between rent-extraction and incentive provision arises.[21]

To be as concise as possible, assume that A is a square matrix with diagonal elements all equal to k, off-diagonal elements all equal $z < k < 1$. Thus, success depends only on whether the action y_a matches the state ω. Let d stand for the default probability that the expert's ex-ante action matches ω without processing s_0. Let h be the probability that s_0 tells the correct state. The scalar $q = (h-d)(k-z)$ stands for the productivity gain from processing information. To save on notation, let the above set of technology parameters be summarized by vector $\underline{q}$.

In this setting, the production of s_0 is *ex-ante contractible*. The correct ex-ante interpretation of s_0, however, cannot be part of an ex-ante contract here. Specifying the correct ex-ante interpretation of s_0 would require not only the agent, but the supervisor, or the principal, to complete task ex-ante.[22]

6.2 Bayesian Monitoring

As the benchmark, the socially optimal care, e_f, is defined by $qp'(e_f) = 1$. Whenever care is not directly observable, there is no guarantee that this first-best is implementable. Rather, if the contract can only be written on the final medical outcome x, the physician's wage is conditioned on this outcome alone. Given the limited liability assumption, the lowest wage the physician can receive is zero. Indeed the optimal liability contract offers such minimal compensation after a failure. The optimal care induced and the efficiency wage paid after a success is determined by the solution to the following problem of the principal[23] :

$$\max_{e, w_S} V_n(r(e, \underline{q}), w) = [p(e)q + dk + (1-d)z](1 - w_S) \tag{6a}$$

$$\text{where } e_n(\underline{q}, w_S) = \arg\max_e [p(e)q + dk + (1-d)z]w_S - e \tag{6b}$$

[21] On the use of limited liability, see the classic work of Sappington (1983) and Innes (1990).

[22] More broadly, the analysis also touches issues relating on the optimal combination of implicit and explicit incentives, e.g. Gibbons (1998), but a careful discussion of this issue, is outside the scope of my paper.

[23] To eliminate unstable equilibria, I assume that $p'''(e) \leq 0$,

Denote the unique solution of the above problem by $e_n(\underline{q})$ and $w_n(\underline{q})$. The next lemma specifies the optimal incentive scheme when examiners observe only the outcome x. This corresponds to a *strict-liability contract*, where the expert is punished after a bad outcome and rewarded after a good one.

Lemma 4 *The optimal care under strict liability is* $e_n(\underline{q}) < e_f(\underline{q})$ *for all* $\underline{q}$.

Suppose now the medical examiner can observe the ex-ante diagnosis and the ex-ante information. As it is well-known since Holmström (1979), obtaining additional reports on the expert's activity allows the principal to construct a more efficient incentive scheme. Given this observation, it is now possible to write *negligence contracts* which rewards the expert if his action matches the ex-ante information, and punishes him otherwise. The scheme ties compensation closer to the unobservable care. To see this, consider the IC condition for the expert;

$$e_m(\underline{q}, w_S) = \arg\max_e p(e)(1-d)w_S + dw_S - e \tag{7}$$

and note that for any given efficiency wage, the radiologist exerts more care than under the strict liability. Formally, for any fixed efficiency wage, the care under negligence is greater than the care under strict liability, i.e., $e_m(\underline{q}, w_S) > e_n(\underline{q}, w_S)$ for all w_S. Since this means that the trade-off between providing incentives and extracting a positive rent from the radiologist is improved, the principal now induces higher care. Let the optimal solution for the principal's problem now be $e_m(\underline{q})$.[24]

Lemma 5 *The optimal care under negligence is* $e_m(\underline{q}) > e_n(\underline{q})$ *and* $EV_m > EV_n$ *for all* $\underline{q}$.

Increasing the observability of the radiologist's activity increases the radiologist incentives to exercise care. In turn, it prompts the principal to induce a higher level of care, and though the care induced is still lower than in the first best, this move from strict liability towards negligence offers the principal a higher surplus from the relationship than before. Closer monitoring combined with a negligence standard is clearly beneficial.

6.3 Biased Evaluations

Let's turn to biased evaluations. In hindsight, a biased evaluator's interpretation of s_0 is distorted towards the true state ω. For simplicity, consider the case of full projection, where the biased interpretation is equivalent to the true state. Information projection induces a type I and a type II

[24] Note that given the observation of $\{y_a, s_0, x\}$, the pair $\{y_a, s_0\}$ is not necessarily a *sufficient statistic* for the agent's effort choice and hence the optimal contract in the $\{y_a, s_0, x\}$ observation case would also condition on x. Considering this possibility does not change the results of this Section.

error in monitoring. A recommendation that was correct ex-ante is mis-classified as an uninformed one if uncertainty resolves in an unfavorable way. A recommendation that was uninformed ex-ante can be classified as a good one if uncertainty resolves in a favorable way.

Importantly, the greater is ex-ante care, the more often mistaken punishments, type I error, happens. The probability of such mistaken negative reports is $p(e)(1-h)$, which is increasing in the level of care and the amount of ex-ante uncertainty. Clearly, if the radiologist takes no care, this mistake never happens. At the other end, if he takes the highest possible care, the probability of this mistake equals the total amount of the ex-ante uncertainty. This means that for any efficiency wage w_S, a physician who anticipates the projection bias, has *lower incentive* to process information carefully. Formally, let his effort choice be $e^{\rho}_m(\underline{q}, w)$, which is now determined by:

$$e^{\rho}_m(\underline{q}, w_S) = \arg\max_e p(e)h(1-d)w_S + dw_S - e \tag{8}$$

Note that for any efficiency wage w_S, the return on care under negligence is diminished relative to the unbiased case.

Proposition 6 *The care under negligence and biased evaluations $e^{\rho}_m(h, w) < e_m(h, w)$ for all w_S. Furthermore, $e_m(h, w)$ is constant and $e^{\rho}_m(h, w)$ is strictly increasing in h.*

The above result has some important corollaries to the use of information in incentive provision. In line with the classic sufficient statistic theorem of Holmström (1979), compensation linked to a performance measure that is sufficient for an alternative one should result in greater efficiency. Under information projection this will no longer be true. A more informative performance measure might create more room for information projection. As a net result, when greater care is expected in the Bayesian case, lower care might result under biased supervisions. Better information under perfect Bayesian processing, will often be more contaminated by the above judgemental errors in my model. A special case when the above result holds is the transition from strict liability to the negligence contract with monitoring. Given Lemma 4, this shift is a Pareto improvement under Bayesian rationality. When uncertainty at the ex-ante stage is sufficiently large, it is a Pareto inferior change when projection bias is present.

Corollary 4 *If $h < h^* \leq 1$ then $e^{\rho}_m(\underline{q}, w_m) < e_n(\underline{q}, w_n) < e_m(\underline{q}, w_m)$ for all $\underline{q}$ and welfare is strictly lower under monitoring than under strict liability.*

Intuitively, under optimal negligence the efficiency wage is lower than under strict liability. This has the effect of decreasing care. At the same time, under negligence the physician's compensation is better linked to how much care he exercises. This increases care relative to the strict liability case. Since better performance information improves the trade-off between rent extraction and incentive provision, the latter effect will always dominate the former in the unbiased case. Under information projection however, a *de jure* negligence rule turns into *de facto* strict liability, hence the wage effect is stronger than the monitoring effect.

It is important to note that similar results will hold even if the efficiency wage w_S is held constant. For example, the supervisor might be able to choose between the following two performance measures: (i) a somewhat noisy performance measure conditioned on the fact whether the event $y_a = s_0$ is true or not, (ii) observing both y_a and s_0. Under Bayesian conditions, (ii) dominates (i) and results in higher care. Typically, the opposite will follow under information projection. More generally, the more a performance measure depends on the separation of ex-ante and ex-post interpretation of the evidence, the less *projection proof* it will be. Hence, the more likely such reversals would occur.

Adopting Bayesian-superior schemes in the above fashion will often backfire, but by accounting for the projection bias of the supervisor, the principal might be able to design a scheme with the same observation that restores the second-best. The next result shows that generically this is not possible. Information projection introduces *noise* in the monitoring process which distorts the trade-off between incentive provision and rent extraction. If the principal correctly anticipates biased supervision, he knows the probability of receiving a wrong report, but not whether the report was actually wrong or right. A positive report might mistakenly reward and a negative report might mistakenly punish. As a result, the return to monitoring is still lower than in the Bayesian case.

To illustrate this mechanism, consider the case, $d = 0$. Here the agent cannot take the ex-post right action without processing s_0 and type II error never occurs. A biased supervisor punishes the agent too often, but efficient incentives can be restored by raising the efficiency wage that ensues positive reports from the supervisor. If $d > 0$, a higher reward also implies giving a higher rent to the agent because rewards are now given not only for exerting effort but also for good luck. This way, the trade-off between incentive provision and rent extraction is distorted. Recognizing this, the principal decides to adopt the monitoring technology less often, and even when he does so, he induces less effort than under Bayesian conditions.

Proposition 7 *Suppose the supervisor's bias is common knowledge between the principal and the agent. The principal induces $e_{bm}(\underline{q})$ and offers wage $w_{bm} = [(h-b)p'(e_{bm})]^{-1}$.*

1 If $d = 0$, then $w_{bm} = \frac{1}{h} w_m$ and $e_{bm}(\underline{q}, w_{bm}) = e_m(q, w_m)$ for all $\underline{q}$.

2 If $d > 0$ and $z = 0$, then $w_{bm} = k w_n$ and $e_{bm}(\underline{q}, w_{bm}) = e_n(\underline{q}, w_n)$ for all $\underline{q}$.

3 If $d, z > 0$*, then* $e_n(\underline{q}, w_n) < e_{bm}(\underline{q}, w_{bm}) < e_m(\underline{q}, w_m)$ *for all* $\underline{q}$*.*

In the above analysis, the projected information led to the exaggeration of the value of s_0 which in effect caused the ex-post exaggeration of the return to effort. It might happen though that the projected information leaves h unaffected, but leads to an exaggeration of d, the probability of taking the right action without processing information. In the above framework, such a mistake does not affect output because it does not change the probability with which an agent is rewarded, conditional on exerting effort. In a more general setup however, where h and d are known ex-ante only to the agent, and where compensation is conditional on the supervisor's assessment of these parameters, exaggerating b might also lower the agent's effort. The results are in stark contrast with the conjecture of the legal literature that the anticipation of hindsight bias prompts agents to exert excess care to avoid ex-post blame, Rachlinski (1998).

Finally, as long as second-best performance measures are not *projection proof*, and they are unlikely to be, the ranking of information sets for efficient incentive provision will be affected. The theory of the second-best, developed under Bayesian assumptions, e.g., Holmström (1979) and Grossman and Hart (1983), would need to consider the judgemental errors that arise due to information projection when considering the design of optimal incentive schemes and the optimal regulation of liability.

7 Exchange of Information

Above I focused on applications where informational differences arise due to the temporal resolution of uncertainty. I now turn to the case where informational differences are present because people receive information from different sources. Communication is that activity whose primary goal is to bridge informational differences. The evidence from Section 2 suggests that information projection plays a significant role in this domain. To establish the first-order implications of my model to this context, I focus on pure coordination problems where the interests of the sender and the receiver are perfectly aligned. Both the sender and the receiver take an action, y_s and y_r respectively, to maximize the probability that the receiver's action matches the state ω. In the first example, limits to communication by the set of messages the sender can use to convey information. In the second example, in the observational learning literature, e.g., Banerjee (1992) and Bikhchandani et al. (1992), similar limits are present when the sender can only recommend an action rather than specify the sources of his information.

7.1 That Obscure Talk of Experts

Efficient information transmission requires a speaker to tailor her messages to the background of her audience. A computer manual that is perfectly informative for an electrical engineer is often meaningless for the average user; the value of a message is determined not by its content alone, but by the relation between its content and what the audience knows and does not know to begin with.

There are three signals in this environment: s_1 the technical jargon or expertise, s_2 a specific technical term, s_3 a noisy lay description. The first two pieces of information are perfect *complements.* The technical term only describes the state when the technical language is known. Formally, the state ω can be decomposed into $\omega = \varpi_1 \varpi_2$ where for simplicity $\varpi_1, \varpi_2 \in \{-1, 1\}$. The sender knows $s_1 = \varpi_1$ and $s_2 = \varpi_2$. The lay description is an imperfect substitute of these two, $\Pr(s_3 = \omega \mid \omega) = h < 1$, whose interpretation is independent of the value of s_1. Communicating the technical language and more than one message is prohibitively costly. Hence the sender can send (i) the technical term, (ii) the lay description, or (iii) remain silent. The cost of sending a message is c.

For simplicity, I assume that the advisor knows the properties of s_3, but not its realization. The insights extend to the case where this assumption is relaxed. The table below summarizes a ρ-biased perception of payoffs corresponding to each of these choices:

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

An unbiased sender chooses the lay description or remains silent. A biased sender deviates from this in three ways. First, if only the lay term was available, she would remain silent too often. Second, if only the expert term was available, she would send this too often rather than stay silent. Finally, when all options are present, she too often prefers the expert term to the lay term. Overall, she *under-communicates* information that substitutes for her expertise, and *over-communicates* information that complements her expertise. Specifically, if the precision of the lay description is high relative to her bias, she communicates too rarely. If the reverse is true, she communicates too often, but sends a message that is too complex.

Proposition 8 *If $\rho < \frac{h-1/2}{1-h}$, the advisor sends the lay term when $(h - \frac{1}{2}) > \frac{c}{1-\rho^2}$, and stays silent otherwise. If $\rho > \frac{h-1/2}{1-h}$, the advisor sends the expert term when $\frac{1}{2}\rho(1-\rho) > c$, and stays silent otherwise.*

Consistent with earlier intuition, experts too often send messages that complement their background information but are too vague for their audiences to understand. In addition, adding communication options can strongly *lower welfare* in the above proposition. Furthermore, removing the option of remaining silent will not necessarily improve welfare if the sender is sufficiently biased, $\rho > \frac{h-1/2}{1-h}$. Simply mandating communication does not restore efficiency. A protocol, which restrict the use of s_2, but encourages or mandates the use of s_3 is necessary. In this manner, publishers should choose people to proof-read a manuscript who does not share the expertise of the author. To improve doctor-patient communication, a hospital should not simply mandate communication, this might backfire, but suggest a language for the interactions. Finally, to facilitate coordination, managers in a hierarchy should not simply be required to provide instructions to subordinates on a set of questions, but may be given specific guidelines on the set of admissible answers to these questions.

The importance of protocols is further enhanced if we consider the attributions the supervisor makes her subordinates upon observing their actions. Since the supervisor exaggerates how well subordinates could have performed if they wished to, she underestimates their attentiveness to her instructions or their loyalty in general. As a response she may terminate the communication relationship institutionalizing the inefficiency. Furthermore, because the supervisor explains differences in the compliance with her advice by differences in subordinates competence rather than the difference in the background information of these subordinates, ϖ_1, she will mistakenly *discriminate* against people with dissimilar backgrounds. Disadvantaged subordinates, such as immigrant workers, anticipating this, will not join the corporate hierarchy at the first place, but decide to choose self-employment instead.[25] Formally, let $\theta \in [0,1]$ be a measure of loyalty or attentiveness which stands for the probability that the subordinates attends to the supervisor's message.

Corollary 5 *Let the subordinate's private information be* $s_0 \in \{-1,1\}$ *such that* $\Pr(s_0 = \varpi_1 \mid \varpi_1) = k$ *for all* ϖ_1*. The supervisor's expected posterior* $E[\pi_1^\rho]$ *is decreasing in* ρ *and increasing in* k *in the sense of fosd for all* π_0*.*

7.2 Credulous Herds

Projection bias also shapes the way privately informed receivers interpret the messages of privately informed senders and it causes a strong form of *imitation.* Specifically, assume that both the advisor (sender) and the advisee (receiver) have private information about the state $\omega \in \Omega$, and share the common objective of maximizing $-(y_r - \omega)^2$. Before taking his action y_r, the advisee again listens to the advisor's advice how can now directly recommend an action y_e. To effectively learn from the recommendation, the advisee has to combine what he learns from y_e with what he privately knows.

[25] The use of communicational protocols may then attracted high-skilled immigrant workers to join corporate hierarchies who fearing underestimation, otherwise might choose self-employment, such as opening a restaurant, instead.

If he projects his private information, however, he exaggerates the extent to which the advisor's action already reflects this information. As a result, he *conforms too much* to the action of the advisor. He fails to adapt the advise to his own circumstances, and instead of learning from it, he imitates it too much.[26]

The above framework is the important context of observational learning with common objectives as introduced by Banerjee (1992) and Bikhchandani et al. (1992). The literature has identified conditions under which coarseness in the action space relative to the signal space causes people to rationally ignoring their private information when learning from the actions of their predecessors. Importantly though, in the absence of such coarseness, these models predict perfect information aggregation. Furthermore, absent mistakes, social learning is always efficient from an ex-ante point of view, and members of a herd have correct confidence in the fact that the action they are taking is the right one. In contrast, the above analysis suggests that a herding motive much stronger under than under rational information processing. Herds will arise due to information projection in many of the financial contexts where rational herding implies perfect information aggregation.[27] Moreover, members of a biased herd will often be *too confident* that the action they are taking is the correct one, and be ex-ante strictly better-off by not knowing what actions their predecessors took

8 Conclusion

Building on extensive evidence, the goal of this paper has been to improve our understanding of how people behave when facing asymmetric information. I offered a widely applicable model that accommodates a variety of social mispredictions and offers novel predictions in a number of domains. Although I focused on applications to agency and communication settings, information projection is likely to matter in a variety of other contexts - bargaining, trade, voting, signalling – as well.

The presence of information projection is robust to various debiasing techniques e.g., Fischhoff(1982), Anderson et al.(1997), Sanna et al.(2002), but further evidence is required to test the predictions of the model. Indeed, it is the articulation of the identifying properties of this model that allows one to specify these predictions and to link behavioral and welfare consequences. Propositions 3 and 4 provide conditions to identify the presence of information projection in individual

[26] Formally, consider that the advisee has to take an action $y_e \in \mathbb{R}$ as close as possible to state $\omega \in (-\infty, \infty)$ on which the shared prior is $N(0, \sigma_0)$. The advisee receives the recommended action of the advisor y_r whose preferences are aligned with his. The advisor's private information before she takes her action is $s_r = \omega + \varepsilon_r$, such that her posterior is $N(\widehat{s}_r, \widehat{\sigma}_r)$. The advisee's private information be $s_e = \omega + \varepsilon_e$ such that her posterior is $N(\widehat{s}_e, \widehat{\sigma}_e)$. If $y_e^{\rho}(y_r)$ is the strategy of a ρ-biased advisee, then one can show that $E\left|y_r - y_e^{\rho}(y_r)\right|$ is decreasing in ρ and $E\left|y_r - y_e^{1}(y_r)\right| = 0$.

[27] See Eyster and Rabin (2009) who offer an extensive analysis of why rational information processing does not produce imitation, rather leads to *waves* and *strong anti-imitation* in most economically relevant social learning environments.

belief-updating. The model can also be tested on disaggregated choice data by combining choices over information sets and choices over economic outcomes, in a manner similar to the design of Loewenstein et al.(2006). The comparative static results allow one to identify the presence and the economic significance of projection bias both in the laboratory and in the field more broadly.

As it is the case for time preferences and risk preferences, there might be heterogeneity in the degree to which different people project various pieces of information. Importantly, my model does not eliminate or explain this heterogeneity, rather it leaves it for further empirical research. As emphasized in Section 3, the identifying property and the key predictions of the model depend solely on the fact that information is projected, $\rho > 0$, relative to the Bayesian null hypothesis that $\rho = 0$.

Ignorance Projection. A direction to extend the ideas presented in this paper is to consider a related phenomenon of *ignorance projection.* In the language of my model, ignorance projection happens when someone who does not observe a signal underestimates the probability with which this signal is available to others. Such a person might understand that others received some information that she did not, but underestimate the extent to which their beliefs will differ from her beliefs. I am not aware on any systematic evidence that would establish ignorance projection, though the major experimental paradigms that establish information projection, e.g., Fisschoff (1975), Newton (1990) do not allow for this possibility.[28] The model offered in Section 3 however provides a tool to identify ignorance projection in the data and also to compare its impact to that of information projection. Importantly, the results of this paper are robust to the presence of potential ignorance projection.

Projection on Future Selves. Although the paper focuses on interpersonal information projection, in Section 3, I mentioned the possibility of extending the model to projecting information onto one's own future selves Such projection might explain *overconfidence in prospective memory*, Erickson (2009). In particular, a person who projects information onto her future selves, will be overconfident about the probability that she will be able to recall her current information in the future. Similarly to the case of studying for an exam with a solution manual, Section 4, here information projection will offer specific predictions on the relation between overconfidence and informational differences. Specifically, a person will have less inflated beliefs about the likelihood that she will memorize the route to her friend's place before looking at the map compared the same belief she forms after looking at the map. This fact has implications to understanding the optimal design of

[28] Importantly, the anchoring assumption employed in the literature to capture the hindsight bias, CLW (1989) would not able to distinguish between *information and ignorance projection* because underestimating mean beliefs is not sensitive to quality of informational differences.

deadlines, reminders and provides an enriched understanding of the role information and memory play in self-control problems.

Informational Differences and Conflict. Recall the dinner example of Section 3. The more the guest knew what the host did not, the lower the guest's assessment was of the kind and the similarity of the host. This example can be extended to understand the role of informational differences in generating intergroup conflict. As people fail to adopt the right perspective when interpreting the actions of others, the presence of informational differences become a direct source of conflict.[29] As groups receive information from increasingly different sources, perceived conflict between these groups will intensify. An analysis of conflict based on information projection, or the combination of information and ignorance projection, will have specific predictions on the set of social learning environments that will exacerbate conflict and social learning environments that will likely reduce conflict.

[29] This problem is often called naive realism in the social psychology literature. For related evidence see the survey of Emily Pronin, Puccio Carolyn, and Lee Ross (2002),

9 Appendix

Proof of Proposition 1. To prove this proposition, I first show that projection shifts probabilistic weight from less to a more informative information sets in the sense of Blackwell (1953). Note that given any $I_i, I_k \subset N$, $I_i \cup I_k$ is always weakly more informative than I_i. Let denote the posterior on ω induced by information I_i by σ_i and analogously σ_{i+k} for $I_i \cup I_k$. Since Ω and Z are finite, σ_i and σ_{i+k} are finite and we can collect their realizations for all possible realizations of signals $\{s_j\}_{j=1}^N$ into matrices Σ and $\widehat{\Sigma}$ respectively. Note that the expectation of σ_{i+k} given I_i is always equal to σ_i. It follows that there exists a Markov-matrix T, i.e., a non-negative matrix with columns summing up to 1, such that $\Sigma = T\widehat{\Sigma}$. The next step of the proof follows from Blackwell (1953).[30] If $\Sigma = T\widehat{\Sigma}$ where T is a Markov-matrix, then for any fixed von-Neumann Morgenstern utility function, $u_i(y, \omega)$, and signal realization it is true that $E[u^*_{I_i}] \leq E[u^*_{I_i \cup I_k}]$ where $u^*_{I_i}$ is defined by Eq.(2).

Given an environment Γ, p_i^0 and p_i^1 induce two pdfs over N such that one can generate p_i^1 from p_i^0 by reallocating probabilistic weight from less informative to more informative information sets. It then follows that $f^1(u_i^*)$ fosd $f^0(u_i^*)$. Since $f^\rho(u_i^*)$ is the probabilistic mixture of $f^1(u_i^*)$ and $f^0(u_i^*)$ where the probabilistic weight shifted is increasing in ρ, the proposition follows. Corollary 1 follows from the fact that projecting a more informative information set in the sense of Blackwell, leads to shifting probabilistic weight to information sets with greater expected utility. .

Proof of Proposition 2. Let the cardinality of the outcome set X be X and let's index its elements by l in ascending order. By the law of conditional probability, $E[f^0(\theta \mid x)] = f(\theta)$ for all priors $f(\theta) \in \Delta\Theta$. Consider the case where $\rho > 0$. Expected belief about θ is given by

$$E[f^\rho(\theta \mid x)] = \sum_{l=1}^{X} f^\rho(\theta \mid x_l) f^0(x_i) = \sum_{l=1}^{X} \frac{f^\rho(x_l \mid \theta) f(\theta)}{f^\rho(x_l)} f^0(x_l) \tag{9}$$

Fixing θ, we can re-write this as

$$E[f^\rho(\theta \mid x)] = f(\theta) \left[\sum_{l=1}^{X} f^\rho(x_l \mid \theta) \frac{f^0(x_l)}{f^\rho(x_l)} \right] = f(\theta)[\lambda^\rho(\theta)] \tag{10}$$

where $\lambda^\rho(\theta)$ is a short-hand for the term in the square-brackets. I now show that $\lambda^\rho(\theta)$ is decreasing in θ for all ρ. Formally, this means that

$$\sum_{i=1}^{X} [f^\rho(x_l \mid \theta) - f^\rho(x_l \mid \theta')] \frac{f^0(x_l)}{f^\rho(x_l)} > 0 \tag{11}$$

Consider first the lowest outcome $x_1 \in X$ and the corresponding term, $[f^\rho(x_1 \mid \theta) - f^\rho(x_1 \mid$

[30] For a proof based on Blackwell (1953) see Hirschleifer and Riley (1999) pp. 193.

$\theta')][f^0(x_1)/f^\rho(x_1)]$. This is negative by virtue of the monotone-likelihood ratio property in θ. Consider now x_1 and x_2 jointly and let's show that

$$[f^\rho(x_1 \mid \theta) - f^\rho(x_1 \mid \theta')]\frac{f^0(x_1)}{f^\rho(x_1)} > [f^\rho(x_2 \mid \theta') - f^\rho(x_2 \mid \theta)]\frac{f^0(x_2)}{f^\rho(x_2)} \tag{12}$$

If $f^\rho(x_2 \mid \theta') < f^\rho(x_2 \mid \theta)$, we are done. If $f^\rho(x_2 \mid \theta') > f^\rho(x_2 \mid \theta)$, then $f^\rho(x_2 \mid \theta') - f^\rho(x_2 \mid \theta) < f^\rho(x_1 \mid \theta) - f^\rho(x_1 \mid \theta')$ must be true by virtue of because of the monotone-likelihood ratio property in θ and because $f(x \mid u, \theta)$ satisfies the monotone-likelihood ratio property in u the inequality follows. Based on a similar logic, it is true that for all $L < X$

$$\sum_{l=1}^{L}[f^\rho(x_l \mid \theta) - f^\rho(x_l \mid \theta')]\frac{f^0(x_l)}{f^\rho(x_l)} > 0 \tag{13}$$

Finally, since $[f^\rho(x_X \mid \theta) - f^\rho(x_X \mid \theta')]\frac{f^0(x_X)}{f^\rho(x_X)} < 0$ Eq.(14) is satisfied for all $\rho > 0$.

Since $\lambda^0(\theta) = 1$ for all $\theta \in \Theta$, for any $\theta^* \in \Theta$,

$$\int_{\theta<\theta^*} f(\theta)\lambda^\rho(\theta)d\theta > \int_{\theta<\theta^*} f(\theta)\lambda^0(\theta)d\theta$$

and thus the result is true when comparing the Bayesian and the biased inference. To show that the same relation holds for $\rho < \rho'$ note that given Proposition 1, $\frac{f^{\rho'}(x_l)}{f^{\rho'}(x_{l'})} > \frac{f^\rho(x_l)}{f^\rho(x_{l'})}$ whenever $l > l'$. This concludes the proof. .

Proof of Proposition 3. The result follows form the above proof with $X = 2$. Note that $\max y' A\sigma_1$ is increasing in θ and in the probability that s_1 is available. If this probability is denoted by p_1, then Eq. (4) gives rise to a stochastic process $\pi(x_S \mid \theta, p_1)$ which satisfies the monotone-likelihood ratio property in θ and p_1 and thus the result follows from the fact that p_1^ρ is increasing in ρ. For Corollary 2, note that $E^{\rho=1}[u_a^*] - E^{\rho=0}[u_a^*] = \int_0^1 Eu_a^*(\sigma_1^1(\theta))\pi_0(\theta)d\theta - \int_0^1 Eu_a^*(\sigma_1^0(\theta))\pi_0(\theta)d\theta$ and then the result again follows from Proposition 2. .

Proof of Proposition 4. By definition, $\pi_1^\rho(\theta \mid x_S)$ fosd $\pi_1^0(\theta \mid x_S)$ when $\int_0^{\theta^*} \pi_1^0(\theta \mid x_S)d\theta \geq \int_0^{\theta^*} \pi_1^\rho(\theta \mid x_S)d\theta$ for all $\theta^* \in [0,1]$. We can re-write this as

$$\pi^\rho(x_S)/\pi^0(x_S) \;\geq\; \left(\int_0^{\theta^*} \pi^\rho(x_S \mid \theta)d\Pi_0(\theta)\right) \;/\; \left(\int_0^{\theta^*} \pi^0(x_S \mid \theta)d\Pi_0(\theta)\right) \text{ for all } \theta^* < 1. \tag{14}$$

where $\pi^\rho(x_S) = \int_0^1 \pi^\rho(x_S \mid \theta)d\Pi_0(\theta)$. The result is thus true given the fact that $\int_0^1 \pi_1^\rho(\theta \mid x_S)d\theta = 1$ for all ρ.

Note finally, that if $\pi^{\rho}(x_S \mid \theta)/\pi^0(x_S \mid \theta)$ is increasing in θ, then $\pi^{\rho}(x_S \mid \theta)/\pi^{\rho'}(x_S \mid \theta)$ is increasing in θ whenever $\rho > \rho'$ and analogously for the decreasing case. To conclude the proof, if s_1 does not require skill to be processed and s_1 complements s_0 then $\pi^{\rho}(x_S \mid \theta)/\pi^0(x_S \mid \theta)$ is increasing in θ for all ρ. Similarly, if s_1 substitutes for s_0, then $\pi^{\rho}(x_S \mid \theta)/\pi^0(x_S \mid \theta)$ is decreasing in θ. .

Proof of Proposition 5. Let $g(s_1, s_0') = \max_{y \in Y} yA\sigma_1(s_0, s_0', s_1) - \max_{y \in Y_e} yA\sigma_1(s_0, s_0')$ and $g(s_1, s_0) = \max_{y \in Y} yA\sigma_1(s_0, s_1) - \max_{y \in Y} yA\sigma_1(s_0)$ where the argument of σ_1 refers to the information that induces is. If s_0' and s_1 are substitutes, $g(s_1, s_0') < g(s_1, s_0)$. If s_0' and s_1 are complements, $g(s_1, s_0') > g(s_1, s_0)$.

Let $E[\pi_1^{\rho} \mid s]$ be the expected assessment when the set of ex-ante produced signals is s. The decision to produce s_0' depends on the sign of the following expression:

$$b(E[\pi_1^{\rho} \mid s_0]) - b(E[\pi_1^{\rho} \mid (s_0 + s_0')]) + a \tag{15}$$

Setting $a = 0$, given Corollary 2, a sufficient condition for the above expression to be positive is that $g(s_1, s_0') > g(s_1, s_0)$. Similarly, when $g(s_1, s_0') < g(s_1, s_0)$, then the above expression is negative for all $\rho > 0$. It follows that the cut-off $a(m, \rho)$ is decreasing in m and if signals are substitutes and increasing in m if they are complements. .

Proof of Lemma 3. First let's derive the optimal contract as given by Eq. (11). The principal's maximization problem yields the following Lagrangian:

$$\mathcal{L}(w_S, e, \mu) = (p(e)q + bk + (1-b)z)(1 - w_S) + \mu(p'(e)qw_S - 1)$$

The FOC with respect to e is given by $p'q(1 - w_S) + \mu p''qw = 0$ and with respect to w_S it is $-(p(e)q + bk + (1-b)z) + \mu p'(e)q = 0$. Solving for μ and substituting for $w_n = 1/p'(e)q$ the equilibrium effort level is given by

$$qp' = 1 - \frac{p''(p + (bk + (1-b)\dot{z})/q)}{(p')^2} = 1 - \frac{p''(p + b/(h-b) + z/q)}{(p')^2} \tag{16}$$

Let the solution of this equation be denoted by $e_n(\underline{q})$. Note that the second-order conditions are satisfied as long as $p'''(e_n(\underline{q})) \leq 0$. An increase in k or h increases q and hence increases the LHS of Eq.(25). An increase in k or h decreases the RHS of Eq.(25). Since p is increasing and concave and $p''' \leq 0$, it follows that this leads to a higher equilibrium effort level. To see the effects of an increase in k and h on the principal's welfare note that for a given w_S, $(p(e)q + bk + (1-b)z)(1 - w_S)$ is increasing in a since $w_S < 1$. Furthermore the optimal w_S given k and h cannot be larger than the original w_S because $h-b<1<p'$ and $k-z<1<p'$. .

Proof of Lemma 4. Let's first derive the optimal contract given monitoring. The principal's maximization problem yields the following Lagrangian:

$$\mathcal{L}(w_S,e,\mu)=(p(e)q+bk+(1-b)z)-(p(e)(1-b)+b)w_S+\mu(p'(e)(1-b)w_S-1)$$

The first-order condition with respect to e is given by $p'q-p'(1-b)w_S+\mu p''(1-b)w_S=0$ and the first-order condition with respect to w_S is given by $-(p(1-b)+b)+\mu p'(1-b)=0$. Solving for μ and substituting $w_m=1/p'(1-b)$ we get that the equilibrium effort level e_m is determined by

$$p'q=1-\frac{p''(p+b/(1-b))}{(p')^2} \tag{17}$$

Note first that $(bk+(1-b)z)/(h-b)(k-z)>b/(1-b) \iff b(k-z)+z(1-b)>bh(k-z)$ which is always true if $h<1$. If we compare Eq. (26) with Eq. (25), it follows then that effort is greater under monitoring because for any e the LHS's of these two equations are the same and the RHS of Eq. (26) is smaller than the RHS of Eq.(25). Given the assumption that $p'''\leq 0$ the result follows.

To show the increase in the principal's welfare note that

$$EV_n=p(e_n)q+bk+(1-b)z-(p(e_n)+b/(h-b)+z/q)/p'(e_n)$$

and

$$EV_m=p(e_m)q+bk+(1-b)z-(p(e_m)+b/(1-b))/p'(e_m)$$

Since $(1-1/p'(e_m))$ and $(1-1/p'(e_m))$ are both positive because $p'(e_n),p'(e_m)>1$, and because $b/(h-b)+z/q>b/(1-b)$ if $h<1$, it follows that $EV_m>EV_n$. .

Proof of Proposition 7. Let's fix a wage w_S. It follows from the discussion in the text that the agent's effort choice is given by the following maximization problem

$$e_m^1(\underline{q},w_S)=\arg\max_e p(e)h(1-b)w_S+bw_S-e \tag{18}$$

The first-order condition is then given by $p'h(1-b)w_S=1$. It is easy to see that for any given w_S, $e_m^1(\underline{q},w_S)<e_m^0(h,w_S)$ as long as $h<1$ and also that $e_m^1(\underline{q},w_S)$ is increasing in h. .

Proof of Corollary 5. To see this corollary note that $e_m^0(\underline{q}, w_S)$ does not depend on h directly. It follows that $e(w_n, h)$ is such that $p'(e) = p'(e_n)$ and $e_m^1(\underline{q}, w_m^0)$ satisfies $p'(e) = p'(e_m^0)/h$. Hence for any $p'(e_n) < \infty$ there exists h^* such that if $h < h^*$ then $p'(e_m)/h > p'(e_n)$. This implies that $e_m^1(\underline{q}, w_m^0) < e_n(\underline{q}, w_n)$. Since social surplus is increasing in e as long as $qp' > 1$ it follows that monitoring decreases social surplus. .

Proof of Proposition 8. To prove this Proposition, we have to consider the principal's problem when the principal knows that the agent's action is given by $e_m^1(\underline{q}, w)$. Here the principal's Lagrangian is given by

$$\mathcal{L}(w_S, e, \mu) = p(e)q + (bk + (1-b)z) - p(e)(h-b)w_S - bw_S + \mu(p'(e)(h-b)w_S - 1)$$

the first-order condition with respect to e is given by $p'q - p'(h-b)w_S + \mu p''(h-b)w_S = 0$ and the first-order condition with respect to w_S is given by $-p(h-b) - b + \mu p'(h-b) = 0$. Solving for μ and substituting $w_{bm} = 1/p'(h-b)$ we get that $e_{bm}(\underline{q})$ is given by:

$$p'q = 1 - \frac{p''(p + b/(h-b))}{(p')^2} \tag{19}$$

Comparing Eq. (28) with Eq.(26) it follows $e_{bm} < e_m$ as long as $h < 1$ because the RHS of (28) is always greater than the RHS of Eq.(26). Furthermore since the RHS of (28) is decreasing in h, $e_{bm}$is decreasing in h. .

References

[1] Anderson, J. C., Jennings M. M., Lowe D. J. and Reckers P. M. J. 1997. "The mitigation of hindsight bias in judges' evaluation of auditor decisions." *Auditing: A Journal of Practice and Theory*, 16(2), 20–39.

[2] Baker George, Robert Gibbons and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, Vol. 109, No. 4 1125-1156

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

[4] Baron-Cohen, Simon, Alan Leslie, and Uta Frith. 1985. "Does the autistic child have a 'theory of mind'?" *Cognition*, Vol. 21, 37–46.

[5] Banerjee, Abhijit. 1992. "A Simple Model of Herd Behaviour." *Quarterly Journal of Economics*, Vol. 107, No.3., 797-817.

[6] Bernstein, Daniel. Cristina Atance, Geoffrey Loftus, and Andrew Meltzoff. 2004. "We Saw It All Along:Visual Hindsight Bias in Children and Adults." *Psychological Science,* Vol. 15, 264 - 267.

[7] Biais, Bruno, and Martin Weber. 2008. "Hindsight bias and investment performance." Mimeo *IDEI Toulouse.*

[8] Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy,* Vol.100, 992-1026.

[9] Birch, Susan and Bloom Paul. 2003. "Children are cursed: An asymmetric bias in mental-state attribution." *Psychological Science* Vol., 14, 283 - 286.

[10] Birch, Susan and Bloom Paul. 2007. "The Curse of Knowledge in Reasoning About False Beliefs." *Psychological Science* Vol. 18, 382 -386.

[11] Berlin, Leonard and Roland Hendrix. 1998. "Perceptual Errors and Negligence." *American Journal of Roentgenology*, Vol 170, 863-867.

[12] Berlin, Leonard. 2000. "Malpractice issues in Radiology: Hindsight bias." *American Journal of Roentgenology*, Vol. 175, 597-601.

[13] Berlin, Leonard. 2004. "Outcome Bias." *American Journal of Roentgenology*, Vol.183, 557-560.

[14] Blackwell David 1953. "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics*, 24, 265-272.

[15] Camerer, Colin. 1987. "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." *American Economic Review.* Vol. 77, No. 5., 981-997.

[16] Camerer, Colin, George Loewenstein, and Martin Weber. 1989. "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy,* Vol. 97, No. 5., 1234-1254.

[17] Camerer, Colin. 1995. "Individual Decision Making." in John Kagel and Alvin Roth eds. *The Handbook of Experimental Economics,* Princeton University Press, Princeton.

[18] Camerer, Colin, and Ulrike Malmendier. 2007. "Behavioral Economics of Organizations" in: P. Diamond and H. Vartiainen (eds.), *Behavioral Economics and Its Applications*, Princeton University Press, Princeton.

[19] Caplan RA, Posner KL, Cheney FW. 1991. "Effect of outcome on physician judgments of appropriateness of care." *Journal of the American Medical Association,* 265, 1957-1960.

[20] Crawford, Vincent, and Nagore Iriberri. 2007. "Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" *Econometrica,* Vol. 75 No. 6., 1721-1770.

[21] Dewatripont, Mathias, Ian Jewitt and Tirole, Jean, 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, Vol. 66, No 1., 183-98.

[22] Durell, Alan. 1999. "Attribution in Performance Evaluation." *Ph.D Thesis Harvard University*, Cambridge MA.

[23] Erickson, Keith. 2009. "Forgetting that we Forget" *Journal of the European Economic Association,* forthcoming.

[24] Esponda, Ignacio. 2008. "Behavioral Equilibrium in Economies with Adverse Selection." *American Economic Review,* Vol. 98 No.4., 1269-91.

[25] Eyster, Erik, and Matthew Rabin. 2005. "Cursed Equilibrium." *Econometrica*, Vol. 73 No. 5., 1623-1672.

[26] Eyster, Erik, and Matthew Rabin. 2009. "Naive Herding." *mimeo* UC Berkeley and LSE.

[27] Fischhoff, Baruch. 1975. "Hindsight / foresight: The effect of outcome knowledge on judgement under uncertainty." *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 1, 288-299.

[28] Fischhoff, Baruch. 1982. "Debiasing," in: D. Kahneman, P. Slovic and A. Tversky (eds.), *Judgement under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

[29] Gibbons, Robert. 1998. "Incentives in Organizations," Journal of Economic Perspectives, Vol. 12, No. 4, 115-32.

[30] Gilovich, Thomas, Kenneth Savitsky, and Victoria 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.

[31] Gilovich Thomas, Victoria Medvec, and Kenneth 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.

[32] Grossman, Sanford and Oliver Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica,* Vol. 51, No. 1, 7-45.

[33] Guilbault, Rebecca, Fred Bryant, Jennifer Brockway, and Emil Posavac. 2004. "A meta-analysis of research on hindsight bias." *Basic and Applied Social Psychology.* Vol. 26, No. 2-3, 103-117.

[34] Harley, Erin. 2007. "Hindsight Bias in Legal Decision Making." *Social Cognition,* Vol. 25, No 1., 48-63.

[35] Heath, Chip, and Nancy Staudenmayer. 2000. "Coordination Neglect: How Lay Theories of Organizing Complicate Coordination in Organizations." *Research in Organizational Behavior*, Vol. 22, 153-191.

[36] Hermalin, Benjamin. 1993. "Managerial Preferences Concerning Risky Projects." *Journal of Law, Economics, & Organization*, Vol. 9, No. 1, 127-135.

[37] Hinds, Pamela. 1999. "The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance" *Journal of Experimental Psychology: Applied* Vol. 5, No. 2., 205-221.

[38] Hirschleifer, Jack and Riley John.1999. *The Analytics of Uncertainty and Information.* Cambridge University Press, Cambridge.

[39] Holmström, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics*, Vol. 13, 324-40.

[40] Holmström, Bengt. 1982. "Managerial Incentive Problems–A Dynamic Perspective." published in Review of Economic Studies. 1999. Vol. 66, No. 1, 169-82.

[41] Inhelder, Barbel and Piaget Jean. 1958. *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures.* New York: Basic Books

[42] Innes, Robert. 1990. "Limited liability and incentive contracting with ex-ante action choices." *Journal of Economic Theory,* Vol. 52, No.1, 45-67.

[43] Jackson, Rene, and Alberto Righi. 2006. *Death of Mammography: How Our Best Defense Against Cancer Is Being Driven to Extinction.* Caveat Press.

[44] Jenter, Dirk, and Fadi Kanaan. 2009. "CEO Turnover and Relative Performance Evaluation" mimeo *Stanford GSB*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=885531

[45] Keysar, Boaz and Ann Henly. 2002. "Speakers' overestimation of their effectiveness." *Psychological Science*, Vol. 13, 207–212.

[46] Jewitt, Ian. 1997. Information and Principal- Agent Problems. *mimeo* Oxford.

[47] Jolls, Christine, Cass Sunstein and Richard Thaler. 2000. "A Behavioral Approach to Law and Economics." in Cass Sunstein ed. *Behavioral Law and Economics*, Cambridge University Press, Cambridge.

[48] Kessler, Daniel, and Mark McClellan.1996. "Do Doctors Practice Defense Medicine?," *Quarterly Journal of Economics,* Vol 111, No. 2., 353-390.

[49] Kessler, Daniel, and Mark McClellan. 2000. "How Liability Law Affects Medical Productivity." *NBER Working Papers* No. 7533.

[50] Kruger, Justin, Epley Nicholas, Jason Parker, and Zhi-Wen Ng. 2005. "Egocentrism over e-mail: Can people communicate as well as they think?" *Journal of Personality and Social Psychology,* Vol. 89, No. 6., 925-936.

[51] Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics,* No. 118, Vol. 4., 1209-1248.

[52] Loewenstein, George, Don Moore, and Roberto Weber. 2006. "Misperceiving the Value of Information in Predicting the Performance of Others." *Experimental Economics,* 3., 281-295.

[53] Milgrom, Paul. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics,* Vol. 12, No 2. 380-391.

[54] Newton, Elisabeth. 1990. "Overconfidence in the Communication of Intent: Heard and Unheard melodies." Unpublished doctoral dissertation, Stanford University, Stanford, CA.

[55] Prendergast, Canice. 1993. "A Theory of Yes Men." *American Economic Review*, Vol. 83, No. 4., 757-770.

[56] Prendergast, Canice, and Robert Topel. 1996. "Favoritism in Organizations." *Journal of Political Economy*, Vol. 104, No. 5., 958-978.

[57] Pronin, Emily, Thomas Gilovich and Lee Ross. 2004. "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self versus Others." *Psychological Review*, Vol. 111, 781-799.

[58] Pronin, Emily, Puccio Carolyn, and Lee Ross. 2002. "Understanding Misunderstanding: Social Psychological Perspectives." in Gilovich, Griffin, and Kahneman eds. *Heuristics and Biases: The Pscychology of Intuitive Judgment.* Cambridge University Press, Cambridge.

[59] Rabin, Matthew, and Dimitri Vayanos. 2009. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies,* forthcoming.

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

[61] Rachlinski, Jeffrey. 1998. "A Positive Psychological Theory of Judging in Hindsight." *The University of Chicago Law Review*, Vol. 65, No. 2., 571-625.

[62] Sanna, Lawrence, Norbert Schwarz and Shevaun Stocker 2002. "When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight." *Journal of Experimental Psychology* Vol. 28, No. 3, 497–502

[63] Sappington, David. 1983. "Limited liability contracts between principal and agent," *Journal of Economic Theory*, Vol. 29, No. 1., 1-21.

[64] Scharfstein, David and Jeremy Stein.1990. "Herd Behavior and Investment." *American Economic Review*, Vol. 80, No. 3., 465-479.

[65] Studdert, David, Michelle Mello, William Sage, Catherine DesRoches, Jordon Peugh, Kinga Zapert, and Troyen Brennan. 2005. "Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment." *Journal of the American Medical Association,* Vol. 293., 2609-2617.

[66] Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under uncertainty: Heuristics and biases." *Science*, 185, 1124-113.

[67] Viscusi, Kip, and Richard Zeckhauser. 2005. "Recollection Bias and the Combat of Terrorism." *Journal of Legal Studies,* Vol. 34, No. 1., 27-55.

[68] Van Boven, Leaf, Gilovich Thomas, and Victoria Medvec. 2003. "The Illusion of Transparency in Negotiations." *Negotiation Journal,* April, 117-131.