Does consistency predict accuracy of beliefs?: Economists surveyed about PSA

Berg, Nathan and Biele, Guido and Gigerenzer, Gerd

2010

Does Consistency Predict Accuracy of Beliefs?: Economists Surveyed About PSA

Nathan Berg
nberg@gmail.com

Guido Biele
biele@mpib-berlin.mpg.de

Gerd Gigerenzer
gigerenzer@mpib-berlin.mpg.de

August 11, 2010

Abstract: *Subjective beliefs and behavior regarding the Prostate Specific Antigen (PSA) test for prostate cancer were surveyed among attendees of the 2006 meeting of the American Economic Association. Logical inconsistency was measured in percentage deviations from a restriction imposed by Bayes' Rule on pairs of conditional beliefs. Economists with inconsistent beliefs tended to be more accurate than average, and consistent Bayesians were substantially less accurate. Within a loss function framework, we look for and cannot find evidence that inconsistent beliefs cause economic losses. Subjective beliefs about cancer risks do not predict PSA testing decisions, but social influences do.*

Keywords: logical consistency, predictive accuracy, elicitation, non-Bayesian, ecological rationality

Corresponding author: Nathan Berg, School of Economic, Political and Policy Sciences, University of Texas-Dallas, 800 W. Campbell Rd., GR31, Richardson, TX 75083-3021.

Does Consistency Predict Accuracy of Beliefs?: Economists Surveyed About PSA

> *For judged probabilities to be considered adequate, or rational, internal consistency is not enough.* –Tversky and Kahneman (1974, p. 1130).
>
> *It appears that a minimal requirement of rationality is that one not hold beliefs that are contrary to objectively available data, coupled with logical, statistical, or mathematical reasoning.* -- Gilboa, Postlewaite and Schmeidler (2009, p. 290)

Section 1: Introduction

We told125 male attendees at the 1996 ASSA meetings in Boston about two widely accepted estimates in the medical literature relating to prostate cancer: the unconditional probability of prostate cancer among asymptomatic men in their 50s, which is 0.025; and the probability (in the same population) of a positive Prostate Specific Antigen (PSA), a commonly used blood test to screen for prostate cancer, which is 0.050. After being informed of these probabilities, we elicited subjective beliefs about two related conditional probabilities: the posterior probability of cancer given a positive PSA test, denoted P(C|+), and the probability of a positive PSA test conditional on undiagnosed prostate cancer, referred to in the medical literature as the *sensitivity* of the PSA test, denoted P(+|C).[1]

Figure 1 summarizes the information provided to subjects about unconditional probabilities of the PSA test and of prostate cancer, P(+) = 0.050 and P(C) = 0.025, and the two conditional beliefs elicited from subjects (with subscripts indexing individuals subjects), $P(C|+)_i$ and $P(+|C)_i$. We invite the reader to pause for a moment of introspection: What numerical values would you assign as your best estimates of P(C|+) and P(+|C)? The novel aspect of this elicitation of conditional beliefs is that it yields a measure of Bayesian consistency without requiring factually accurate beliefs. Elicited conditional beliefs can be completely wrong, yet entirely consistent with the definition of conditional probability. If people vary in the extent to which they adhere to Bayes' Rule, then would we expect this to correlate with other observable features?

[1] Although the medical literature refers to the posterior probability of cancer conditional on a positive test result as the test's *positive predictive value*, this paper follows convention in economics referring to P(C|+) as the *posterior probability.*

A wide range of voices has remarked upon the centrality and singularity of Bayes' Rule as both a prescriptive and descriptive norm.[2] Gilboa, Samuelson and Schmeidler (2010, p. 1), for example, write: "The mode of reasoning most widely used in economic modeling is *Bayesian*." Starmer (2000, p. 377) writes that, before non-additive probability models appeared in the economics literature, economists usually took it for granted (and probably continue to take for granted) that the Savage Axioms—which guarantee that choice over lotteries can be represented as expected utility maximization with respect to a subjective probability distribution conforming to Bayes' Rule—provide the "right model of individual choice." Reinhardt Selten (2001, p. 13) writes that "Modern mainstream economic theory is largely based on an unrealistic picture of human decision making [in which] agents are portrayed as fully rational Bayesian maximizers of subjective utility." Camerer et al.'s (2003, p. 1214-1215) definition of "full rationality" requires that "people have well-formed beliefs about how uncertainty will resolve itself, and when new information becomes available, they update their beliefs using Bayes's law." According to Aragones et al., (2005, p. 1364), "Most of the formal literature in economic theory and in related fields is based on the Bayesian model of information processing." And Gilboa, Postlewaite and Schmeidler (2009, p.287) emphasize the singularity of Bayesian information processing (as opposed to a plural toolkit containing multiple procedures for reasoning on the basis of data or lack of data), observing that: "[W]ithin economic theory the Bayesian approach is the sole claimant to the throne of rationality."[3]

[2] Savage argued for a normative interpretation of expected utility theory while admitting that he himself violated the theory when first encountering the pairs of gambles used in Allais' paradox (Savage, 1954). See Starmer (2000, 2009) for more on normative interpretations of expected utility theory.

[3] Binmore (2008) distinguishes Bayesians (i.e., users of Bayesian models in their appropriate context—what Savage described as Small Worlds—where all states and probabilities are known and genuine surprises therefore cannot occur) from "Bayesianismists" (i.e., those who mis-apply Bayesian models built for Small Worlds to Large-World domains, where Binmore and Savage would view it as preposterous to summarize one's thinking by means of a single probability distribution or prior). Gintis (forthcoming) allows that many important decisions may not have well-specified state spaces or well-defined probabilities, which he says calls for extensions of the Bayesian model to those challenging contexts, but with the Bayesian model serving as the singular benchmark model of information processing. Gintis (forthcoming, p. 2) writes: "I have always been comfortable with identifying rationality with the Savage axioms, which may be described in shorthand as 'preference consistency over lotteries with subjective probabilities.'" And Loewenstein (2006) usefully cautions that theoretical extensions of standard models in pursuit of added realism,

Based on this near methodological consensus regarding the centrality of Bayes, we define *consistency of beliefs* as the extent to which subjective conditional beliefs adhere to Bayes' Rule. Because Bayes' Rule is equivalent to the definition of conditional probability, it imposes the following restriction on individuals' subjective conditional beliefs (assuming that probabilities we supplied coincide with subjects' unconditional beliefs):

$$P(C|+)_i \, P(+) = P(+|C)_i \, P(C).$$

The ratio of numerical values for the two unconditional probabilities in the expression above (if probabilistic logic is to be applied to subjective beliefs) requires that the ratio of elicited conditional beliefs takes on a specific numerical ratio:

$$P(C|+)_i / P(+|C)_i = P(C)/P(+) = 0.025/0.050 = ½.$$

One can then measure inconsistency in various ways based on deviations from this restriction.

We define *inconsistency* as the absolute (log approximated) percentage deviation of an individual's elicited ratio of conditional beliefs from the correct ratio of unconditional probabilities:[4]

$$\textit{inconsistency}_i = |\log(\ [\ P(C|+)_i / P(+|C)_i\ ] / [1/2]\ )|.$$

Of 125 respondents who provided a complete set of elicited belief data, 24 (19 percent) generated perfectly Bayesian conditional beliefs, indicated by $\textit{inconsistency}_i = 0$.[5]

Published point estimates for these conditional probabilities are $P(C|+) = 0.34$ and $P(+|C) = 0.68$. Note that one's beliefs can be substantially inaccurate, even as a perfect Bayesian. For example, six perfect Bayesians in our sample reported $P(C|+)_i = 0.50$ and $P(+|C)_i = 1.00$; two reported 0.20 and

especially concerning information and information processing, do not necessarily wind up being more realistic.

[4] The log-approximated percentage deviation from Bayes' Rule has two main advantages over other measures of deviation. First, it attenuates and therefore reduces the influence of extreme deviations, which makes the results we report conservative. Second, unlike exact percentage deviations, the log-approximation is completely symmetric, because $\log(\ [\ P(C|+)_i / P(+|C)_i\ ] / [1/2]\ ) = -\ \log(\ [\ P(+|C)_i / P(C|+)_i\ ] / 2\ )$, and therefore does not depend on whether the restriction is expressed as $P(C|+)_i / P(+|C)_i = ½$ or as $P(+|C)_i / P(C|+)_i = 2$.

[5] Our survey team intercepted ASSA attendees just outside the hall where the main registration desk was, using a scripted 3- to 10-minute face-to-face interview protocol. ASSA attendees were surprisingly agreeable to provide us with subjective beliefs about prostate cancer risks and self-reports about PSA testing. We collected 133 surveys. Eight respondents supplied partial belief data by non-responding to at least one of the five belief items, leaving 125 complete observations. The eight partial responders are excluded from the analysis except where noted otherwise.

0.40; and one reported 0.10 and 0.20, all of which are well into the upper half of the inaccuracy distribution, despite adhering perfectly to Bayes' Rule. There is a problem, however, using the same pair of elicited beliefs to compute both inconsistency and inaccuracy, because those measures are then functionally and statistically dependent. Instead, we use a different, but related, set of survey items to measure accuracy: two subjective beliefs about lifetime (as opposed to point-in-time) risk of prostate cancer, and lifetime probability of mortality from prostate cancer. These beliefs measures of lifetime incidence and mortality depend on roughly the same factual sources, but are numerically very different because the majority of prostate cancers are slow-growing and non-lethal. Computing inconsistency based on the conditional beliefs described earlier and inaccuracy based on lifetime incidence and mortality, we can inspect bivariate covariation in these two variables in the scatter plot shown in Figure 2, with inconsistency on the x-axis versus inaccuracy on the y-axis.

The 24 observations clustered along the y-axis are perfect Bayesians with zero inconsistency. Notice that the two most inaccurate observations are perfect Bayesians. In the other direction, the two most inconsistent observations are well below the midpoint of the inaccuracy range. The distribution consists of a relatively small number of extreme responses which are highly inconsistent and/or inaccurate, and a larger group that is minimally to moderately inconsistent and inaccurate. Overall pairwise correlation is −0.04 and statistically insignificant. Translated into elasticity of inaccuracy with respect to inconsistency, the coefficient from a bivariate regression of accuracy on inconsistency is −0.06 (i.e., elasticity, since both variables are in log units) with t statistic −0.46. Eliminating extreme observations in all combinations that we tried (e.g., throwing away the five largest observations of inconsistency and inaccuracy, or the 10 extremes of both) raises the magnitude of the negative correlation, often dramatically so. For example, if we throw away observations with inconsistency greater than 1.5, pairwise correlation becomes −0.30 with elasticity −0.60 (t statistic = −3.5). That would imply (by linear extrapolation) that beliefs twice as inconsistent as average are expected to be 60 percent more accurate. There is no evidence for positive association between consistency and accuracy.

We speculate that many of us who teach choice under uncertainty might expect (or wish) that different normative metrics (i.e., consistency and accuracy of beliefs) would correlate positively, implying convergence or harmonization among potentially contradictory normative criteria. Suppose, for example, people with fewer transitivity violations (another normative metric based solely on internal consistency) also turned out to be more Bayesian, with more accurate beliefs, higher levels of accumulated wealth, substantially longer lives, superior health, and higher than average levels of self-reported happiness. Then axiomatic rationality based solely on internal consistency might be regarded as standing on a firm evidential basis, bolstering these axioms' intuitive appeal by correlating positively with normative measures that do not depend on internal consistency.[6] Any positive association between consistency and accuracy of beliefs remains, as yet, empirically unsubstantiated as far as we are aware, and is refuted by our data.

As the quotations appearing before the introduction of this article suggest, many of us—when it really matters (e.g., giving advice to a loved one, or a high-stakes medical decision)—apply normative criteria that go beyond, and sometimes contradict, internal consistency.[7] The first issue this paper

[6] The psychologists Hastie and Rasinski (1986) were the first to classify the two distinct categories of normative measures that Gilboa (forthcoming) also discusses (unfortunately, using the same term "coherence" in a manner diametrically opposed to Hastie and Rasinski's usage). Hastie and Rasinski (1986) and Hammond (1996) refer to norms based on internal consistency as *coherence norms* (e.g., Bayesian beliefs, transitivity, Kolmogorov axioms) to distinguish them from non-consistency-based normative metrics based instead on free-standing scales measuring a level of performance, referred to as *correspondence norms*. Correspondence norms are so named because they measure how well an individual's choices or inferences correspond to the demands of his or her environment. The key difference is that correspondence norms (i.e., free-standing level-of-performance norms, which include accuracy of beliefs, accumulated wealth, lifespan, and happiness) can rank the single acts of two people, whereas consistency or coherence norms say nothing about single acts when considered in isolation and only impose restrictions on pairs or larger sets of decisions. Gilboa (forthcoming) poses the question of whether non-consistency-based normative measures such as happiness belong in definitions of economic rationality at all. He makes a strong case for explicitly defining rationality, perhaps pluralistically and with context dependence to bring in criteria other than consistency.

[7] According to an anecdote from reliable sources concerning a well-known proponent of axiomatic decision theory, when faced with the decision of whether to take a job offer from a competing university, the proponent deliberately chose to deviate from the normative theory which he knew well. It was not due to indifference. It was a high stakes decision, and he therefore brought in normative criteria other than consistency to assess what it would mean to make a good decision. When colleagues asked him why he didn't just choose a prior, add up probability-weighted utilities associated with each of his options, and choose according to the criterion of maximum expected utility, the decision theorist replied in exasperation: "Come on, this is serious!" (Gigerenzer, 2004, p. 62). This anecdote illustrates that even those who

seeks to address is how to document empirical regularities linking consistency to the objective accuracy of subjective beliefs. The elicitation technique reported here provides a tool that allows for virtually any functional relationship between consistency and accuracy (measured at the individual level), enabling us to pose the following question as a hypothesis test: "Do people with consistent beliefs also tend to have accurate beliefs?"

The second question concerns whether inconsistency is associated with economic losses. Google Scholar returns more than 4,000 hits associated with the phrase "non-Bayesian beliefs." EconLit returns more than 3,800 hits. Judging from this intense scrutiny by economic researchers, one might presume that deviations from Bayes' Rule have important economic consequences. And they might. Yet one finds little evidence to substantiate the hypothesis of economic losses due to inconsistent beliefs in this same literature.[8] Raising questions about whether deviations from standard normative benchmarks are individually or socially costly (or perhaps even beneficial) should not imply broader skepticism about the substantial experimental evidence documenting anomalies and biases. On the contrary, when one takes the behavioral economics literature seriously, especially its priority on empirical realism, it suggests a much needed follow-up question: If individuals do not conform to standard normative decision-making models, what then is the economic cost?

In search of evidence for direct costs due to inconsistency, we adopt a model which assumes that PSA decisions are based on minimization of a loss function that depends on beliefs about prostate cancer risks, beliefs about the quality of the screening instrument, and the PSA decision itself. We follow Gaechter, Orzen, Renner, and Starmer (2009) in attempting to take advantage of the high level

---

best understand the mechanics of Bayesian reasoning find it inapplicable as a tool, or procedure, for making decisions.

[8] Behavioral economists have paid close attention to modeling deviations from Bayes' Rule, and experimental economists have spent considerable effort documenting the degree to which subjects conform to or deviate from Bayes' Rule (e.g., Camerer 1987, 1992; Ganguly, Kagel and Moser, 2000; Kluger and Wyatt, 2003). The unstated presumption in much of this literature is that people ought to be Bayesian, a point of view that Gilboa, Postlewaite and Schmeidler (2009, p. 286) explicitly challenge, with the observation that an arbitrarily chosen prior in conflict with frequency data would seem hardly rational: "A paradigm of rational belief should allow a distinction between assessments that are well-founded and those that are arbitrary."

of statistical fluency and familiarity with axioms of rational choice among economists by studying data collected from them. Minimization of expected losses leads to an objective function that can, in theory, be influenced by inconsistent beliefs through two distinct channels. The first channel through which losses could occur would be if inconsistency causes inaccurate beliefs, in which case we would expect to find a strong positive association between inconsistency and inaccuracy. The second channel for inconsistency to cause losses would be if inconsistent people had a different likelihood of having a PSA test (net of the effect of inconsistency on subjective beliefs about risks and benefits). In this case, we would expect to estimate a large effect of inconsistency on the PSA decision itself (either positive or negative, since we make no assumption about whether PSA testing is good or bad) in the presence of controls measuring subjective beliefs about cancer risks, benefits and costs of PSA testing. The intuition is simple: if inconsistent beliefs are costly, then one expects that inconsistent men either have less accurate beliefs, or a systematically different behavioral mapping from perceived costs and benefits into PSA decisions.

The third issue addressed in this paper concerns the actual decision process men use to make decisions about getting tested for PSA. Subjective beliefs about cancer risks, the quality of the PSA test, and chances of negative side effects (conditional on surgical or radiation treatment) surprisingly have no predictive power for self-reported PSA decisions. This corroborates what respondents self-reported about their decision-making processes: low rates of search for statistical information and low rates of "weighing pros and cons" even among respondents who identified both benefits and harms. We find that respondents condition PSA testing decisions on social cues (variables coding whom one talked to prior to deciding to get tested). Given strong incentives for doctors to practice defensive medicine, over-test, and over-diagnose (Studdert et al., 2005), it is surprising that economists, who are well aware of incentive-mismatch problems, appear to ignore advice from the National Cancer Institute to weigh pros and cons before testing.

Because there is room for misunderstanding, we want to state explicitly that our goal was not to demonstrate that economists fail to conform to Bayes' Rule. As mentioned, 24 out of 125 conformed perfectly to Bayes' Rule. We want to stress that, in the absence of evidence showing that deviations from Bayes' Rule adversely affect payoffs, we do not interpret these deviations as irrationality. Rather, our goal is to provide an empirically grounded account of the actual decision process that statistically sophisticated decision makers use, revealing what—if any—role internal consistency of beliefs plays.

Section 2 describes how the data were collected and reports descriptive statistics. Section 3 presents the main findings in the form of regressions linking consistency to accuracy and, second, consistency to self-reported PSA decisions. Section 4 investigates the robustness of these findings, presenting further evidence regarding the role of social influences in PSA testing decisions. Finally, Section 5 discusses interpretations of the results and prospects for new norms of rationality that allow for inconsistency.

Section 2: Description of data

Descriptive Data About Survey Respondents

We surveyed attendees of the annual meeting of the American Economic Association (regularly attended by approximately 9,000 registered conference participants), also known as the Allied Social Science Associations meetings, January 6-8, 2006, in Boston, Massachusetts. Our interviewer conducted face-to-face interviews based on a scripted protocol designed to last three to 10 minutes, although no time limit was imposed. The script (reproduced verbatim in Appendix 1) was visible to respondents, and the interviewer encouraged respondents to read any sample items for themselves if they wanted clarification. Most interviews were collected a few meters from the registration desk at the AEA meetings, which also served as a passageway to and from conference sessions. The location was chosen to ensure, as much as possible, representative chances of intercepting different types of conference attendees.

The interviewer approached men only, and only those who appeared to be at least 40 years old. He approached potential survey respondents with a memorized introductory statement offering respondents a choice of $3 cash or a Swiss chocolate bar, and assurances that the survey would be short. Survey respondents who chose $3 instead of the chocolate bar (83 versus 17 percent) were asked if they wanted to donate the $3 participation fee to a cancer charity, which a majority did. Table 1 contains summary statistics for survey responses.

Of 133 respondents, 123 (92 percent) said they were economists. The 10 non-economists described themselves as political scientists or academics working in fields that overlap with economics. A few additional survey items not summarized in Table 1 were collected as well. For example, respondents' subfields revealed a nicely heterogeneous representation of the economics profession, and these subfield indicators are used as controls in some of the regressions reported in the next section. The age distribution was remarkably symmetric, with a mean of 51, and covering a large range, 26 to 79. For the most part, our interviewer succeeded at hitting the over-40 target, with 119 reporting ages of 40 or older.

Nearly half the respondents (46 percent) reported having had a PSA. Among respondents 50 and older, the rate of PSA testing was 65 percent. Most respondents (91 percent of the 124 who responded) said they recommend that men in their 50s have a PSA, with almost no difference in rates of recommendation by age.

Non-Response

In Table 1, the column under the heading(s) "Number of Responses" shows that item-specific non-response was a problem for several questions, although not the ones we would have expected. Nine refused to classify their work as either "more applied" or "more theoretical." No one refused to say whether he had taken a PSA. Nine refused, however, to make a recommendation about whether men in their 50s should have a PSA.

Perceived Harms, Risks and Benefits of PSA Testing

We will return to the remaining items in Table 1 shortly. Before getting to those, Table 2 summarizes eight frequently cited medical studies about the risks and benefits of PSA testing, with comments highlighting statistical findings and expert opinion, especially potential harms from screening, which patients undertaking cost-benefit calculations would likely want to consider. After gaining FDA approval in 1986 for use among men already diagnosed with prostate cancer, PSA testing spread rapidly as a screening tool for asymptomatic men, with some estimating that by the late 1990s as many as half of American men over the age of 50 had undergone PSA testing (Gann, 1997). Aside from the large direct costs of financing mass screening, which have been estimated at $12 to 18 billion per year (U.S. Preventive Services Task Force, 2002, p. 128), another key point of contention regarding PSA screening concerns the benefit of early detection. Most prostate cancers grow so slowly that patients with prostate cancer die of other causes first (Stanford et al., 1999; U.S. Preventive Services Task Force, 2002). The benefits of early detection may also be limited in the case of fast-growing cancers for which treatment has very limited success. While some studies report evidence that early detection of prostate cancer reduces disease-specific mortality, there is no evidence showing reduction in overall mortality (Ciatto et al., 2000; Holmberg, et al., 2002; Yao and Lu-Yao, 2002; Draisma et al., 2003; Concato et al., 2006). The most recent randomized trial in the U.S. found no evidence that PSA screening reduces death from prostate cancer or death from cancer in general; in fact the death rates were slightly higher in the screening group (Andriole et al., 2009). At the same time, the medical literature reports significant harms from prostate cancer screening, including psychological stress, needless biopsies following false positives, and overtreatment of nonlethal prostate cancers that result in complications such as incontinence and impotence (Wang and Arnold, 2002; Hawkes, 2006).

Returning to Table 1, the survey item labeled “Harms?” encodes responses to the forced-choice (yes/no) question: “In your opinion are there potential harms associated with PSA screening?” In light of the medical literature summarized in Table 2, it surprised us that only a quarter of respondents said there were harms associated with PSA testing. Perhaps most surprising was that only about a third of respondents reported weighing pros and cons when deciding whether to have a PSA test. Not weighing pros and cons can, of course, be rationalized if someone perceives zero costs or zero benefits, because in that case there are no tradeoffs to weigh. When it comes to PSA testing, however, the material in Table 2 shows a medical literature that has, from the mid-1990s, emphatically recommended weighing costs and benefits as opposed to automatic screening for asymptomatic patients. We worried, in fact, that this sample item asking whether economists had weighed the pros and cons might not generate any variation, with nearly all respondents answering "Yes."

Elicited Frequencies

The following five probabilistic beliefs were elicited:

- *lifetime incidence* (the probability that a randomly drawn male in the U.S. is diagnosed with prostate cancer within his lifetime) denoted P(C Lifetime)
- *lifetime mortality* (the probability that a randomly drawn male in the U.S. dies of prostate cancer within his lifetime) denoted P(D Lifetime)
- *incontinence probability* (the probability of incontinence conditional on surgical treatment for prostate cancer) denoted P(Incontinence | Surgery)
- *posterior probability* (the probability that an asymptomatic U.S. male in his 50s has prostate cancer conditional on a positive PSA test) denoted P(C|+)
- *sensitivity* (the probability that an asymptomatic U.S. male in his 50s has a positive PSA test conditional on the event that he has prostate cancer at the time of screening) denoted P(+|C).

The bottom block of elicited belief and published point estimates in Table 1 shows that respondents' beliefs about these probabilities tended to be slightly too large, but not far off from published point estimates in the medical literature. Insofar as cost-benefit considerations drive PSA decisions, one would expect these five belief variables to have joint explanatory power as a proxy for perceived net benefits of testing. We test that hypothesis below.

Consistency and Accuracy of Beliefs

We sought to construct a measure of logical inconsistency that does not depend directly on the accuracy of stated beliefs, so that functionally independent measures of consistency and accuracy, based on separate sets of survey items, could be computed for each individual. To accomplish this, the elicitation scheme (as described in the Introduction) allowed for infinitely many pairs of subjective beliefs to be perfectly Bayesian, regardless of accuracy. Our interview script reads:

> The main focus of the survey is prostate cancer and PSA (Prostate Specific Antigen) screening. I won't ask any personal questions about the illness itself, just about screening. I'd like to elicit your best guesses about the risks of prostate cancer.
>
> [*Elicitation of* $P(C\ Lifetime)_i$:] For a randomly drawn American male, I'd like you to guess the probability that he will *be diagnosed* with prostate cancer in his lifetime?
>
> [*Elicitation of* $P(D\ Lifetime)_i$:] What would you say is the probability that he will die from prostate cancer in his lifetime?
>
> Now I'm going to ask you about American males in their 50s who have no symptoms, have never been diagnosed with prostate cancer, and are screened with a PSA test for the very first time. One leading study suggests that 5% of randomly sampled men from this population have a positive PSA. It's also estimated that 2.5% actually have prostate cancer at the time of screening, which includes those whose PSAs failed to detect the disease. [source: Harris and Lohr, 2002, Ann Intern Med].
>
> [*Elicitation of* $P(C|+)_i$:] Given a positive PSA, I'd like you to estimate the probability that a man actually has prostate cancer.
>
> [*Elicitation of* $P(+|C)_i$:] And given cancer at the time of screening, what would you say is the

probability of a positive PSA?

The first two elicited beliefs, lifetime incidence and mortality, are used to construct a measure of belief inaccuracy. The conditional point-in-time beliefs are used to construct a measure of belief *inconsistency*. As is clear from the interview script, applying the definition of conditional probability to these conditional beliefs imposes the restriction: $P(C|+)_i/P(+|C)_i=1/2$. Respondents might know nothing about relevant medical studies and published PSA facts but nevertheless conform perfectly to this restriction and be perfectly Bayesian. Absolute log-approximated percentage deviations from this Bayesian restriction generates our measure of an individual's inconsistency:[9]

$$inconsistency_i = \left| \log[P(C|+)_i/P(+|C)_i] - \log[1/2] \right|.$$

Inaccuracy of beliefs with respect to published point estimates is defined as:

$$inaccuracy_i = (| \log[P(C \text{ Lifetime})_i/0.177)] | + | \log[P(D \text{ Lifetime})_i/0.028] |)/2.$$

This definition computes inaccuracy by averaging absolute percentage deviations of lifetime incidence and lifetime mortality from their respective point estimates in the medical literature.[10] The scatter plot of *inconsistency* and inaccuracy presented earlier revealed zero or negative correlation. If one supposes there is a single scale of general intelligence, or a single-dimensional spectrum of axiomatic rationality, as is commonly implied by references to "rational" versus "irrational" subjects in the behavioral

9 We re-ran regressions reported in the next section with alternative definitions of inconsistency based on other functional specifications of the deviation. For example, deviation can be measured in percentage points as opposed to percentage deviations with the following formula (although it gives disproportionate influence to respondents with large-magnitude beliefs): *absolute percentage-point deviation from Bayes* = $| P(+|C)_i - 2P(C|+)_i |$. Another deviation measure we tried was raw percentage deviations rather than log approximations: *absolute percentage deviation from Bayes* = $| [P(C|+)_i/P(+|C)_i - 1/2]/(1/2) |$, which produces a much more spread out distribution and, unfortunately, is not invariant under algebraically equivalent re-statements of the restriction. For example, a slightly different deviation emerges from the formula: *alternative absolute percentage deviation from Bayes* = $| [P(+|C)_i/P(C|+)_i - 2]/2 |$. The log approximations we use attenuate extreme deviators and produce more conservative (i.e., smaller magnitude) correlations. We also tried binary classification schemes splitting the sample into subsamples of consistent Bayesians versus inconsistent non-Bayesians. Dichotomization as shown in the next table actually strengthens the case for our interpretations.

10 Lifetime incidence and lifetime mortality are used because the point-in-time PSA-related frequencies (*posterior probability* and *sensitivity*) were already used to compute inconsistency. Most of the variation in *inaccuracy* as defined above derives from beliefs about mortality since it is much rarer and therefore generates a wider range of percentage deviations. We re-ran all empirical models using alternative measures of *inconsistency*: lifetime incidence deviations alone, lifetime mortality devaitions alone, and an average of five deviations based on all five elicited beliefs, revealing no substantive changes. Appendices 2 and 3 describe the distributions of these measures in greater detail.

economics literature, one would expect performance according to one normative metric to correlate positively with performance as measured by other normative metrics. These data provide no support for such a theory.

Accuracy and Consistency Within Subsamples

Next, four cuts of the sample are used to divide respondents into consistent and inconsistent subsamples and contrasts in mean inaccuracy are reported. Groupings into consistent versus inconsistent subsamples are shown as columns in Table 3: perfect Bayesians versus deviators from Bayes; below- versus above-median *inconsistency*; bottom versus upper quartiles of *inconsistency*; and Ballpark Bayesians (a very inclusive classification for anyone whose inconsistencies can be modeled as Bayesian beliefs plus a noise term) versus Emersonians (those who commit gross errors in conditional probabilistic reasoning described in detail below).[11] The first column contains mean values of inaccuracy, signed inaccuracy, four log deviations of elicited beliefs, inconsistency and signed inconsistency. Reading horizontally across the first row, Table 3 indicates the average among the 24 perfect Bayesians (those with $inconsistency_i=0$) had higher inaccuracy than the rest of the sample (1.26 versus 0.90). Similarly, the lower half of the inconsistency distribution had higher inaccuracy than the upper half (1.08 versus 0.87), and the lower quartile had higher inaccuracy than the upper quartile (1.26 versus 0.77). According to the fourth cut of the sample into Ballpark Bayesians and Emersonians, accuracy is, once again, negatively associated with consistency (inaccuracy of 1.08 among the consistent versus 0.78 among the inconsistent).

The second row of Table 3 shows that beliefs of consistent respondents tend to be too small, whereas the beliefs of inconsistent individuals tend to overshoot the estimates in medical journals. Consistent individuals' beliefs are not, however, generally any closer to those published estimates.

[11] The label is inspired by Emerson's (1841) "Self Reliance" in which he wrote: "The other terror that scares us from self-trust is our consistency … A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines. With consistency, a great soul has simply nothing to do."

Rows 3 and 4 show log deviations for lifetime incidence and mortality, the two components averaged in *signed* inaccuracy and inaccuracy.

Of the 16 *t* statistics in the middle block of Table 3 labeled under the heading "log deviations," five have magnitude greater than 2, indicating statistically significant unconditional differences in means between consistent and inconsistent subsamples. Of these five significant differences, consistent individuals' mean deviation from zero is smaller in three cases[12] and larger in two.[13] These disaggregated bivariate contrasts, while mixed, do not show any tendency for consistent individuals to have more accurate beliefs, and are generally consistent with the initial view of the bivariate relationship in Figure 2.

Taxonomy of Inconsistencies: Emersonians and Ballpark Bayesians

Closer examination of the elicitation scheme reveals that there are conceptually distinct ways in which a respondent can deviate from Bayes' Rule. Some respondents are within plausible bounds (defined just below) and could be modeled as if they were producing Bayesian beliefs with an error term that produces moderately inconsistent conditional beliefs. Other subjects' beliefs involve more basic violations of inequalities required by conditional probability. The former group is referred to as Ballpark Bayesians and the gross violators of the definition of conditional probability are referred to as Emersonians.

We define three types of gross violations of probability theory, any one of which would indicate a process for generating beliefs that cannot possibly be reconciled with the definition of conditional probability. The first gross logical error is $P(C|+)_i > 0.50$. The definition of conditional probability

[12] The three cases in Table 2 where consistent individuals are, on average, closer to zero deviation than inconsistent individuals are: -0.00 versus 0.22 for log(posterior/0.34) among perfect Bayesians and deviators from Bayes, with *t* statistic -2.1; -0.18 versus 0.48 for log(mortality/0.028) among Ballpark Bayesians and Emersonians, with *t* statistic -2.5; and -0.11 versus 0.67 for log(posterior/0.34) among Ballpark Bayesians and Emersonians, with *t* statistic -7.9.

[13] The two cases in Table 2 where consistent individuals are, on average, farther away from zero deviation are: -0.69 versus 0.23 for log(mortality/0.028) among perfect Bayesians and deviators from Bayes, with *t* statistic -2.2; and 0.13 versus -0.10 for log(sensitivity/0.64) among lower and upper quartiles of the inconsistency distribution, with *t* statistic of 2.5.

states that $P(C|+) = P(C \cap +)/P(+)$. The numerator refers to an intersection of events for which it must be true that $P(C \cap +) \leq \min\{P(C), P(+)\}=0.025$. The unconditional probabilities provided to respondents imply that conditional beliefs must be bounded above by ½:

$$P(C|+)_i \leq 0.025/0.05 = 0.50$$

Elicited probabilities precisely at the upper bound of 0.50 correspond to the belief that there are no false positives. Of 133 respondents, 36 (34 economists and 2 non-economists) violated this logical bound with subjective *posterior* beliefs strictly greater than 0.50.

The second gross departure from probabilistic logic is $P(C|+)_i > P(+|C)_i$. Substituting the definition of conditional probability for both terms, the numerators of the conditional probabilities are of course the same while the denominators take on known values. But $P(C)=0.025 < P(+PSA)=0.05$ implies $P(C|+) \leq P(+|C)$, which holds with equality only when the intersection event in the numerator has probability zero. Eleven respondents strictly violated this condition, 9 of whom also committed the first gross departure.

The third logical error is $P(C|+)_i = P(+|C)_i$. Given the information provided which explicitly mentions false positives and cancers undetected by PSA testing, $P(C|+PSA)$ cannot be zero. The argument in the preceding paragraph implies the sharp restriction $P(C|+) < P(+|C)$. Sixteen respondents provided equal conditional beliefs. Of these, seven also violated the first logical restriction by stating $P(C|+)_i = P(+|C)_i>0.50$. Seven others stated $P(C|+)_i = P(+PSA|C)_i=0.50$. In total, 45 respondents committed at least one of the three errors resulting in the designation *Emersonian*.

Section 3: Evidence that Inconsistency Leads to Economic Losses?

Loss Function With Two Channels For Inconsistency

Denote respondent i's *inconsistency* as δ_i. His probabilistic beliefs, which are a function of inconsistency, are represented as $b_i(\delta_i)$. Person-specific value judgments needed to rank contingent outcomes associated with prostate cancer, PSA testing, and treatment options, are summarized by the parameter vector θ_i, which is interpreted as accounting for all inter-personal differences aside from inconsistency and beliefs. We suppose that ω represents states of nature drawn from a standard probability measure on the universe Ω. In the PSA testing context, states can be thought of as terminal nodes on a large probabilistic event tree generating contingencies that describe various combinations of events: cancer, positive PSA, types of prostate cancer, and treatment options in the event of cancer.

States ω include contingencies with successful treatments, partially successful treatments with side-effects, and contingencies with unnecessary surgeries (i.e., surgery that removes slow-growing cancers which would not have proved lethal if left untreated)—as well as the opposite, contingencies in which valuable treatment options are missed. The first step from the root of the event tree has two branches corresponding to the unobserved events of prostate cancer and no prostate cancer. The second step has four branches total, two branches from the cancer node, and two from the no cancer node, corresponding to the observed events +PSA and -PSA.

The path along the tree corresponding to the joint event "No cancer and -PSA" is a terminal node. We can normalize the payoff associated with this node to zero, indicating a status-quo outcome that abstracts from small monetary, time and hassle costs associated with having the PSA test and receiving a correct, negative result. Along the branch with no cancer and +PSA (i.e., a false positive), several contingencies are possible corresponding to various options given to patients who have a positive PSA. These include watchful waiting (with the stress of worrying about as-yet undiagnosed prostate cancer); biopsy, false positive on biopsy, unnecessary surgery; biopsy, false negative on biopsy, undiagnosed cancer; biopsy, correct positive indicating cancer; or a biopsy that successfully rules out prostate cancer.

From this long yet far-from-complete list of states ω, one appreciates that many person-specific value judgments must go into assigning payoffs so that all contingencies can be ranked. For example, as authors of one of the medical journals quoted in Table 2 wrote, some men will prefer to live fewer years with a fully functioning body, and others will prefer to live more years with side effects of treatment. Still others might prefer to never be tested or diagnosed, regardless of the underlying physical state. The person-specific parameters in θ_i imply that loss functions for different people will take on different values (representing different rankings of states) even if their subjective beliefs are identical: even if $b_i = b_{i'}$ and $\delta_i = \delta_{i'}$, then i and i' will nevertheless assign different losses to each contingency ω, and possibly make different *ex ante* loss-minimizing testing decisions $t_i^* \neq t_{i'}^*$, whenever $\theta_i \neq \theta_{i'}$. This allows for full heterogeneity in ranking the contingent outcomes and does not presume there is a universally correct decision (to test, or not to test).

Given these definitions, the very standard probabilistic structure generating ω conditional on subjective beliefs b_i is summarized by a conditional pdf: $f_{\omega|b}(\omega, b_i(\delta_i))$. The loss function depends on states, inconsistency (which imparts a direct effect on the losses assigned to all contingencies net of its effect on beliefs about the probabilities of reaching any particular node on the tree), the testing decision denoted t_i, and person-specific parameters needed to rank contingencies: $L(\omega, \delta_i, t_i; \theta_i)$. Taking beliefs and inconsistency as fixed, the decision maker computes risk (i.e., expected loss) at each element in the binary choice set (either $t_i = 0$ which codes the decision not to have a PSA test, or $t_i = 1$, which codes the decision to have a PSA):

$$R_0 = \int_\Omega L(\omega, \delta_i, 0; \theta_i)\, f_{\omega|b}(\omega, b_i(\delta_i))\, d\omega,$$

$$R_1 = \int_\Omega L(\omega, \delta_i, 1; \theta_i)\, f_{\omega|b}(\omega, b_i(\delta_i))\, d\omega.$$

Finally, the optimal choice of t_i minimizes risk: $t_i^* = \text{argmin}_{t \text{ in } \{0, 1\}} R_t$.

One sees from this that there are two channels through which δ_i exerts an influence on PSA decisions t_i^* and therefore two channels through which one might observe evidence, albeit indirectly, that δ_i is associated with economically meaningful losses. The empirical strategy is to examine the channels separately (after linearizing the functional dependence of b_i on δ_i and of t_i^* on δ_i). We seek to measure the effect of δ_i on b_i and the effect of δ_i on t_i*. If *inconsistency* leads to losses, then we expect to observe empirical effects of appreciable size through at least one of these two channels.

Chanel 1: Empirical Model With Conditional Effect of Inconsistency on Inaccuracy

Table 4 shows results from a regression of inaccuracy on inconsistency, with controls for whether respondents consulted written information, the mode in which information was processed, social influencers, a quadratic function of age, and subfield indicators along with other personal characteristics from the survey. Comparing the simple bivariate model (in which the regression coefficient on inconsistency was -.06) to the kitchen sink model in Table 4 (in which the coefficient on inconsistency is nearly the same, changing only to -.08), one sees no evidence that inconsistency exerts large conditional effects on the accuracy of beliefs. Similarly for every intermediate specification involving different subsets of the regressors: we never saw a statistically significant and positive coefficient that would demonstrate a positive association between consistency and accuracy of beliefs.

There are several puzzling effects in Table 4 to note, however. Consulting written information paradoxically increases inaccuracy of beliefs. On the other hand, deliberation captured by the variable "weighing pros and cons" appears to have a beneficial effect reducing inaccuracy, with a magnitude just large enough to cancel out the effect of consulting written information. Although 29 respondents report consulting written information and 46 report weighing pros and cons, only 15 do both. Six respondents report having consulted an authoritative source such as a medical journal, which also implies having consulted a written source. The average neoclassical economist and average econometrician were about one third less inaccurate than the sample average.

Chanel 2: Empirical Models With Conditional Effects of Inconsistency on PSA Testing

Table 5 presents estimates of four linear probability models, with *t* statistics computed using robust standard errors.[14] The first three models are the main focus—prediction of PSA decisions. The

[14] Logit and probit models produce qualitatively identical results and are available from the authors upon request. Similar

fundamental model assumes PSA decisions are a function of all five subjective beliefs and a quadratic function of age. The *add-info-processing model* assumes that PSA decisions are a function of everything appearing in the *fundamental model* and, in addition, depend on information acquisition, information processing, and inconsistency. Finally, the *add influencers model* allows the probability of taking a PSA to encompass the two previous models and, in addition, depend on social influencers. The final columns of Table 5 provide a comparison of the same encompassing model applied to a different dependent variable, respondents' PSA recommendations.

We find statistical confirmation of the self-reports that most economists do not weigh costs and benefits in the results of a joint test that the first five regressors have zero coefficients. This corresponds to the hypothesis that subjective beliefs about cancer risks and benefits of treatment do not influence PSA decisions. The second-to-last row of Table 5 shows p-values for that hypothesis, which reveal surprisingly weak predictive power of subjective beliefs in the first two models. This weak predictive power does not result from overall weakness of the prediction equation, however, as likelihood ratio tests easily reject the hypothesis that all coefficients in the model are zero, across all models. According to the p-value in the third model, however, subjective costs and benefits begin to have statistically significant predictive power once information about social influences is added to the model. Even in the *add influencers model*, individual beliefs have surprisingly weak effects on the probability of having PSA testing. For example, the perceived risk of incontinence, which we would have guessed would strongly condition men's evaluations of the test's desirability, has very moderate effects across the three PSA-decision models, implying that a man whose perceived risk of incontinence to be twice as big as average is, at most, 6 to 8 percentage points less likely to have a PSA. Coefficients on information acquisition and processing (i.e., pros-cons deliberation and logical

to Wisdom, Downs and Loewenstein's (2010) approach, we use the linear probability model estimated by OLS (with robust standard errors) to provide easy-to-interpret magnitudes of estimated effects on binary outcomes (healthy versus unhealthy menu choices, in their case, and PSA decisions in ours). The linear probability model has the advantage of easily correcting for heteroscedasticity of errors. We checked that none of the important effect sizes or qualitative results change with logit or probit specifications of the empirical model.

inconsistency) are nowhere statistically significant.

The *doctor influenced* variable reveals strong conditional correlation between reliance on a doctor's recommendation and PSA test taking, despite the obvious incentive mismatch in doctor-patient transactions that lead to well-documented problems of defensive medicine, over-diagnosis, over-prescription, over-treatment and other potential problems that economists should be well aware of (see Behrens, Güth, Kliemt and Levati, 2005; Loewenstein, 2005; and Sorum et al., 2004, for more on doctor-patient incentive mismatch).

Statistical Predictors of the PSA Recommendation?

The simple correlation between PSA recommendations and self-reported decisions is a surprisingly small 0.09 (and far from statistical significance). The last columns of Table 5 show the estimated prediction model applied to PSA recommendations. To keep the sample the same, the PSA recommendation was modified to a forced-choice version that codes non-responses as zeros. Even in this forced-choice version, the rate of recommendation remains nearly twice as large as the rate of PSA taking, 85 versus 46 percent. Beliefs about costs and benefits have more predictive power for PSA recommendations than for PSA decisions but, once again, consistency of beliefs plays a very limited role.

Theories Regarding Inconsistent Beliefs and Other Forms of Inconsistency

Why might smart people hold inconsistent subjective beliefs? Gilboa, Postlewaite, and Schmeidler (2008) provide examples of decision contexts (e.g., wars, or a coin that one has never seen or flipped before) in which they argue it would be irrational to hold probabilistic beliefs. Non-standard reasoning processes that generate behavior inconsistent with axioms of internal consistency can be defended and, in some contexts, shown to enjoy advantages over decision processes adhering strictly to consistency (e.g., Gilboa and Schmeidler, 1995; Samuelson, 2001; Aragones et al., 2005; Spiegel, Heifetza and Shannon, 2007; Robson and Samuelson, 2009). Grunwald and Halpern (2004) identify the problem of dilation—where updating newly arrived information can cause posterior distributions to become more spread out and therefore less precise—to argue that non-Bayesian updating which sometimes ignores information provides more precise predictions. This less-is-more result regarding the number of variables used in prediction tasks appears in a growing number of theoretical and empirical studies (e.g., Hogarth and Karelia, 2005, 2006; Baucells, Carrasco and Hogarth, 2008; Berg and Hoffrage, 2008; Goldstein and Gigerenzer, 2009). The finding that less information can enhance performance also appears in laboratory studies (Camerer, Loewenstein and Weber, 1989) and financial data (DeMiguel, Garlappi, and Uppal, 2009).

In a related vein, models of time-inconsistency (Loewenstein, 1987) and the possibly adaptive advantages of time-inconsistency have been discussed (Halpern, 1997; Robson and Samuelson, 2009; Warneryd, forthcoming). One empirical study showed that time-inconsistency and expected utility violations were both associated with higher payoffs, inside and outside the task domain that generated those inconsistencies (Berg, Eckel and Johnson, 2010). Theoretical work on rule-based behavior typically considered to be incompatible with axiomatic rationality has stimulated discussions about inconsistencies that provide compensating benefits of simplicity and robustness in the face of Large-World uncertainty (Gigerenzer and Selten, 2001; Bewley, 2002; Segal and Sobel, 2007; Comte and Postlewaite, 2008; and see also Gintis, 2010, on Homo Ludens).

Normative Status of Bayesian Reasoning, Money Pumps and Dutch Books

The Savage axioms underlying expected utility theory are a prime example of consistency criteria whose normative status is widely accepted despite a lack of evidence demonstrating that deviators suffer significant losses. Sugden (1991) argues (with great originality in the face of near methodological consensus pointing in the opposite direction) against the normative interpretation of expected utility theory. Hammond's (1998) model formalizes the argument made informally many times before advocating a strong normative interpretation for expected utility theory and the Bayesian mechanism that supports it. Starmer (2000, 2005, 2009) provides truly illuminating historical and methodological analysis of normative debates about Bayesian reasoning and expected utility theory.

Similar to Bayesian consistency, preference consistency is assumed in virtually every model with utility functions and often defended as normatively appealing based on inconsistent agents' theoretical vulnerability to money-pump or Dutch Books exploitation (Davidson, McKinsey and Suppes 1955, p. 146; Raiffa 1968, p. 78).[15] Although the existence of transitivity violations is by now beyond doubt (Tversky, 1969; Grether and Plott, 1979; Loomes, Starmer and Sugden, 1989, 1991; Sippel, 1997; Harbaugh, Krause and Berry, 2001; Andreoni and Miller, 2002; List and Millimet, 2004), there seems to be little evidence that individuals who behave inconsistently in real economic environments suffer significant losses as a result. Chu and Chu (1990) and Cherry, Crocker and Shogren (2003) report some instances of individuals who are money-pumped in the lab, showing that they quickly learn to avoid inconsistent choices that leave them vulnerable to exploitation. List and Millimet (2004) show that subjects in the field vary significantly in terms of consistency of choice patterns, and that market experience reduces the probability of inconsistent patterns of choice—without showing, however, that inconsistency leads to reduced levels of economic performance. Although experience or contact with market pricing mechanisms can reduce buy/sell disparities and facilitate efficient pricing, such experience does not necessarily make individual-level inconsistencies disappear (Loomes, Starmer, and Sugden, 2010) and sometimes is associated with new inconsistencies (Braga, Humphrey, and Starmer, 2009). Camerer and Hogarth (1999) suggest that learning about the consequences of one's inconsistency occurs relatively slowly, and Loewenstein (1999, 2005) argues that many high-stakes decisions, especially medical decisions, are one-shot—without repetition in the

15 Exceptions include a growing number of papers, some previously mentioned, including the models of Rubinstein and Spiegler (2008), Laibson and Yariv, (2007), Cubitt and Sugden (2001), and DeLong, Shleifer, Summers and Waldman (1991), in which inconsistent individuals do not necessarily succumb to exploitative competitors. Fehr and Tyran (2005) and Halitwanger and Waldmand (1985) emphasize the role of strategic complementarities in determining whether inconsistency among a few individuals influences aggregate measures of economic performance, while Sen (1993) argues against the normative appeal of internal consistency axioms in general.

decision maker's natural environment—raising questions about whether economists should assume that inconsistency is likely to be exploited in competition and therefore mitigated by experience. Rubinstein and Spiegler (2008) critique money pump arguments on the grounds that actually carrying out exploitative transactions requires face-to-face contact that very likely triggers an attitude of caution or suspicion among the potentially exploitable. As Rubinstein and Spiegler (2008, p. 237) put it, "We tend to think strategically about the situation and suspect that there is a 'catch,' even if we cannot pinpoint it."

Section 4: Decision Making Process in PSA Testing

As mentioned earlier, only 46 out of 128 respondents reported having weighed pros and cons when deciding on PSA testing. Among those who did not weigh pros and cons were 16 who did not despite having reported that they perceive harms. This clear departure from thought processes typically assumed in economics motivates us to look for more evidence about the decision making process. The importance of modeling thought processes rather than restricting analysis to outcomes or consequences motivates the admittedly speculative considerations that follow and attempts to cull additional information from our data (Tukey, 1977; Rubinstein and Osborne, 1988; Leland, 1994; Gigerenzer and Selten, 2001; Bardsley et al., 2010). This section attempts to follow Rubinstein's recommendation (Rubinstein,1991, 2001, 2003, 2006) to open the "black box" of decision processes in more detail. Decision processes other than cost/benefit calculus can perform well by various normative metrics and, as numerous evolutionary models have shown, can be rationalizable under mechanisms that generate selective pressure.[16]

One of the most frequently encountered non-standard decision procedures in evolutionary

[16] Rubinstein argues that experiments yield most insight when examining assumptions and documenting regularities that provide an evidential basis for others to inductively generate new theories (rather than testing predictions of theory). Sugden (2008a, 2009), too, points to a role for reporting on empirical regularities that can be used as an evidential basis for others to inductively construct new theory. Binmore et al. (2002) reflect on the importance of investigating more deeply into how people think about games. Gilboa, Postlewaite and Schmeidler (2009, p. 285) advocate "a view of rationality that requires a compromise between internal coherence and justification…."

models is imitation and, more generally, conditioning action on social cues (see Boyd and Richerson, 1985, or Gintis, forthcoming, and the references therein regarding imitation and adaptive success). One important source of justification in the social and family environments in which we make medical decisions is social cues. This section attempts to provide additional insight into the role of deliberative reasoning, the search for information, and social cues.

Table 6 presents a cross-tabulation of responses to the harms question, and the pros and cons question. Non-responses are recorded, too, because they might contain information about decision processes. To examine whether the joint distribution of harms and weighing pros and cons is any different among PSA takers, Table 6 indicates in brackets the number within each cell who are self-reported PSA takers. The joint distributions among PSA-takers and non-PSA-takers are remarkably similar. The respondents in the diagonal elements of the bivariate distribution include 59 respondents who do not see harms, which provides a plausible rationalization for having not weighed pros and cons. Twelve respondents' reports were entirely consistent with cost/benefit calculus. Respondents in the off-diagonal positions are, however, more difficult to square with cost-benefit calculus, raising the question of how they are choosing to get tested, if not by a process of weighing pros and cons?

Guess-50 Heuristic

One possibility is that, with no incentive payments for accurate guesses or (more likely, we think) reflecting honestly on their ignorance about statistical facts of PSA and prostate cancer, respondents simply guess 50 (as a default belief based on the standard appeal to symmetry). It costs very little effort if it is the default belief about binary outcomes in the absence of data. We coded the number of times respondents guessed "50 percent" to see if completely uninformed priors, or use of a *guess-50 heuristic,* was correlated with *consistency* or *accuracy*. Among the five elicited beliefs about probabilities, the maximum number of times anyone in the sample guessed 50 is twice. Interestingly, those who guessed 50 twice had more accurate beliefs, with mean inaccuracy of 0.71 (sd 0.01) among the 22 respondents who guessed 50 twice, versus 1.02 (sd 0.09) among those who never guessed 50. Of the 24 perfect Bayesians, two guessed 50 twice. Emersonians and Ballpark Bayesians guessed 50 at roughly the same rate. And inconsistency was uncorrelated with guessing 50. Appendix 4 discusses a negative finding—no natural frequency effect—relating to evidence in the psychology literature that communicating probabilities in natural frequencies (e.g., "7 in 1000" instead of "0.7 percent") can lead to dramatic improvements in Bayesian reasoning and significantly different medical decisions.

Additional Evidence Regarding Social Influences on PSA Decisions

The paired rows of Table 7 present mean contrasts between subsamples that correspond to different hypotheses about the role of particular variables in influencing PSA decisions. The first pair of rows shows the main finding, which is a large difference in the rate of PSA taking between those who reported nobody influenced them and those who reported at least one influencer (36 versus 78 percent). No other variable has such a large bivariate association with PSA taking. The remaining pairs look for other variables and interactions that modulate the effect of social influence.

The second pair of rows in Table 7 looks for an effect of weighing pros and cons among those who reported being influenced by at least one other (most likely, a spouse). In this subsample of socially influenced respondents, rates of PSA testing show virtually no effect from weighing pros and cons. The third pair of rows in Table 7 shows the difference in rates of PSA testing among those who weigh pros and cons and those who do not, revealing a modest 15 percentage point difference: 76 versus 61 percent. As with all bivariate contrasts, causality is of course unclear. One explanation for higher rates of PSA testing among those who weighed pros and cons is that, after getting tested as a result of a social heuristic, these respondents then gathered information and weighed pros and cons as an after-the-fact rationalization.

The fourth pair of rows in Table 7 casts some doubt on what exactly those who report weighing pros and cons are weighing. Among those who weigh pros and cons, there is only a slight difference in rates of PSA testing between those who perceive harms and those who perceive no harms: 86 and 76 percent, respectively. Similarly, the sixth pair of contrasts shows that among those who perceive harms, those who weighed pros and cons and those who did not have similar rates of PSA testing, although the small number of observations makes these comparisons imprecise.

After social influences, the second largest bivariate contrast was between those who consulted written sources and those who did not (the 3rd from the bottom pair in Table 7), with rates of PSA testing of 95 and 55 percent, respectively. While this could have occurred as the result of information search consistent and subsequent weighing of costs and benefits according to the standard model, we strongly doubt it. Much of the research literature on PSA testing in recent years has reported proven harms and no proven benefits associated with screening asymptomatic populations. We would have guessed that reading the medical literature would lead economists to greater skepticism about the benefits of PSA testing. For example, the sources in Table 2 caution that discovering more cancers and discovering them earlier does not imply saving lives. Another interesting statistical issue in prostate cancer risk studies is that PSA testing was shown to reduce disease-specific mortality but not overall mortality. If weighing pros and cons caused the PSA decision rather than the other way around, then the difference in rates of PSA taking within pros-and-cons weighers should be especially large between those who perceive harms and those who do not (which it is not).

One reading of these data is that those who perceived harms felt a greater need to rationalize their decision to get tested by reporting that their testing decisions resulted from a systematic process of weighing pros and cons. This is consistent with the fourth through last rows of Table 7. The next-to-last (seventh) pair in Table 7 is consistent with this hypothesis of after-the-fact rationalization: if consulting information led to higher rates of testing based on information discovered in those sources, then it would presumably matter whether one weighed those factors or decided in some other way. The seventh pair shows that, among those who consulted written sources, there is nearly the same rate of

PSA testing between those who weighed pros and cons (9 out of 9) and those who did not (10 out of 11). Finally, the last pair of rows in Table 7 indicates that consistency of beliefs, once again, plays a very limited role in explaining PSA decisions, revealed by similar rates of PSA taking among perfect Bayesians and extreme deviators (i.e., Emersonians).

Section 5: Discussion

Findings

The first objective was to elicit belief data in way that would provide independent measures of consistency (with respect to Bayes' Rule) and the objective accuracy of subjective beliefs. Our elicitation technique gives respondents two unconditional probabilities and then elicits related conditional probabilities to accomplish this objective. The second objective was to document evidence consistent with economic losses due to inconsistent beliefs. The data we collected revealed no positive correlation between consistency and accuracy, implying that inconsistent beliefs did not generate economic losses by reducing the accuracy of beliefs, at least in the context we studied. The other channel capable of signaling the economic losses would have been a strong conditional effect of inconsistency on the probability of getting a PSA test, which we also did not find. Finally, we estimated a linear probability model of men's decisions about PSA testing and found that subjective beliefs about risks, benefits and costs are jointly non-predictive. Just about any variable that belongs in a standard expected utility model failed to predict PSA decisions. However, once information about social influences was added to the empirical model, the subjective beliefs became jointly statistically significant, and the model's sign pattern became amenable to straightforward interpretations.

With full awareness of the usual caveats needed in interpreting self-reports about issues as personal as medical decision making, we asked respondents how much written information they had acquired, the sources of that information, and whether or not they had weighed pros and cons in deciding whether to have a PSA test. More than half said they had not weighed pros and cons. Insofar as the standard information processing model provides a poor fit of the data, one may rightfully ask whether these data are simply too noisy to reveal real underlying statistical links. We argue, on the contrary, that respondents' self-reported PSA decisions become intelligible, with acceptable levels of model fit, under the alternative hypothesis that economists, like many people, sometimes rely on a simple heuristic of following doctors' advice—especially when sitting in a hospital or doctor's office—which could be referred to as a *white-coat heuristic*: *See a white coat, do what it says.* The social influencer indicator variables, especially *doctor influenced*, add significant predictive power. Whether trusting one's doctor is effective in any normative sense is not addressed by our findings.

Why Economists?

To improve the chances of finding empirical links between logical consistency and objective accuracy of beliefs, the data reported in this paper were collected mostly from economists. Gaechter, Orzen, Renner, and Starmer (2009) argue that empirical findings of anomalous behavior in samples of economists are especially convincing, since one would expect economists' professional training to sensitize them to mechanisms causing these effects. Presumably the self-awareness of economists makes anomalous effect sizes smaller than in the general population and therefore those effects can be interpreted as conservative lower bounds. Our sample size of 133 was comparable to theirs, which was120. Previous studies have shown that economists behave differently from non-economists because of both selection and training (Carter and Irons, 1991; Frank, Gilovich and Regan, 1993; Yezer, Goldfarb and Poppen, 1996). Surveys of economists have also shown that economists' statistical reasoning and policy views differ substantially from those of non-economists, even after controlling for education, income and gender (Caplan, 2001, 2002; Blendon et al., 1997). Also relevant to the medical decision-making data studied in this paper is previous survey evidence showing that economists agree more than non-economists on the determinants of health and healthcare expenditures (Fuchs, Krueger and Porterba, 1998). Perhaps the most compelling reason for studying economists is that their beliefs about statistical and medical concepts can be measured with far less noise than in the general population, whose poor understanding of statistics and "health literacy" is well documented (Williams et al., 1995; Baker et al., 1998; Parker et al., 1995; Lusardi and Mitchell, 2009).

Logical consistency undoubtedly enjoys objective normative status in particular task settings, for example, when taking the GRE exam. However, a growing body of theoretical models suggests that deviations from standard normative axioms in economics, surprisingly perhaps, may have beneficial effects for individual and aggregate welfare.[17] Historians of science have also pointed out that willingness to hold inconsistent views is a regularity rather than an exception among innovators, for example, Kitcher (1992, p.85), who writes:

> [O]n numerous occasions in the history of science, investigators have found themselves inclined to accept the members of a set of statements that they could recognize as jointly inconsistent, without knowing immediately what should be abandoned: Darwinian evolutionary theory survived Lord Kelvin's estimates of the age of the earth, Bohr's theory of the atom was retained and developed even though it was at odds with classical electromagnetic theory. The phenomenon should be apparent from humbler situations, in which people know that they are inconsistent but do not yet see the right way to achieve consistency. It may even be universal, if each of us is modest enough to believe that one of our beliefs is false."

The conclusions we draw are not categorically against the real-world benefits of adhering to axioms of logical consistency. Rather, our goal is to emphasize the importance of matching normative

---

[17] There is also a growing literature concerning benefits of inaccurate (distinguished here from inconsistent) beliefs. Complementing psychological studies of so-called self-serving bias, Samuelson and Swinkels (1996) report advantages in learning for those with distorted beliefs. Inflated beliefs about the value of one's endowment can increase payoffs in bargaining (Dekel and Scotchmer, 1999; Heifetz and Spiegel, 2001; Heifetz and Segev, 2004). Having a reputation for being illogical in financial markets can make it difficult for opponents to predict one's actions (Kyle and Wang, 1997). And overconfidence in the advice of financial experts can increase market liquidity, resulting in equilibria with distorted beliefs that Pareto-dominate rational expectations (Berg and Lien, 2005; Berg and Gigerenzer, 2007). Recently, Gilboa and Samuelson (2010) study a learning environment in which biased minds learn more effectively.

criteria to particular decision-making contexts, while providing a counterexample in which standard normative benchmarks are violated and performance is unchanged (if not improved). Economists, who are presumably as familiar with the normative benchmarks as anyone, vary substantially in the degree to which they conform to consistency benchmarks, in the accuracy of their beliefs, and in the medical decisions they make. And yet, statistical links between these different sources of variation are mostly weak. Descriptively, social influences appear to be at least an order of magnitude more important than the fundamentals of perceived risks and benefits of PSA screening.

A Bolder Normative Economics in Which Inconsistency Is Allowed?[18]

Our first finding, that consistency does not predict accuracy, suggests that the usual notions of axiomatic or consistency-based rationality are poor proxies for context-specific notions of rationality, sometimes referred to as ecological rationality (Gigerenzer and Selten, 2001; Smith, 2003). The second finding that consistency is uncorrelated with actual decision outcomes (when taken together with the first) suggests that inconsistency in this domain has a small economic cost. The third finding, that social influences are necessary to make sense of the empirical PSA decision model, reveals the importance of social cues. Conditioning action on social cues no doubt functions well in many contexts, but is surprising in light of well-known incentive problems in doctor-patient transactions.

Rubinstein (2006) expresses doubt that economic theory, normative or descriptive, serves the prescriptive function that many, if not most, economists have in mind when defending policy implications based on economic research. Gintis (2010), while arguing for the centrality of the rational actor model, allows that it will be necessary and desirable to pursue extensions of standard notions of rationality in contexts that take us outside the small worlds to which the Bayesian model is applicable. With a slightly different take on the same theme, Gilboa (forthcoming) writes in support of pluralistic approaches rather than the *one-axiom-fits-all-contexts* approach to normative analysis, which is

[18] See Berg and Gigerenzer (2010) on narrow normative interpretations of rationality axioms in behavioral economics.

prevalent if not dominant in both neoclassical and behavioral economics. More explicitly, Gilboa, Postlewaite and Schmeidler (2009, p. 288) write:

> We reject the view that rationality is a clear-cut, binary notion that can be defined by a simple set of rules or axioms. There are various ingredients to rational choice. Some are of internal coherence, as captured by Savage's axioms. Others have to do with external coherence with data and scientific reasoning. The question we should ask is not whether a particular decision is rational or not, but rather, whether a particular decision is more rational than another. And we should be prepared to have conflicts between the different demands of rationality. When such conflicts arise, compromises are called for. Sometimes we may relax our demands of internal consistency; at other times we may lower our standards of justifications for choices. But the quest for a single set of rules that will universally define the rational choice is misguided.

Tversky and Kahneman (1986) argued for a research program that maintains strict separation between normative and descriptive analysis, arranged in a clear hierarchy, with normative on top. Contemporary behavioral economics has enthusiastically undertaken this program whose ground rules hold that no descriptive finding is allowed to raise doubts about the normative authority of neoclassical rationality axioms. Thaler (1991) had already taken up this program in 1991, going to great pains to reassure unconvinced readers that behavioral economics posed no threat to neoclassical norms and, in fact, had nothing to add to normative economics since it had already reached a state of perfection enjoying broad consensus among economists (Berg, 2003). Tversky and Kahneman (1986), in the conclusion of their article, suggest a role for policy to help those who deviate from the normative model to conform. The notion that decision models should serve as tools for aiding real-world decisions is one that Rubinstein (2001, p. 618) rejects: "To draw an analogy, I do not believe that the study of formal logic can help people become `more logical', and I am not aware of any evidence showing that the study of probability theory significantly improves people's ability to think in probabilistic terms."

Some behavioral economists and their colleagues (e.g., Jolls, Sunstein and Thaler, 1998) invest a degree of faith in the prescriptive value of neoclassical rationality axioms that one rarely finds in the neoclassical literature, with calls for interventions to "de-bias" those of us who deviate from axiomatic rationality. Behavioral economists' frequent empirical investigation of "biases" and "deviations" from norms of rationality—expected utility violations, preference reversals, time inconsistency, and non-Nash play in laboratory games—seems to harden the normative authority of neoclassical models. These models may be descriptively wrong, the thinking goes, but they nevertheless provide the reliable guidance about what people ought to do.

Ariely, Loewenstein and Prelec (2003) show that many of the predictable properties of aggregate demand curves based on standard consumer theory need not be abandoned as empirical regularities, despite strong evidence refuting the axiomatic assumptions of the underlying model of consumer choice. Even those of us whose policy preferences are influenced by the rich contributions of economic theory that motivates a role for government (e.g., based on externalities, market power, and information asymmetries) can enthusiastically join Libertarian critics such as Sugden (2008b), whose article titled, "Why incoherent preferences do not justify paternalism," says it all. He is, like we are, methodologically committed to challenging axiomatic rationality, which lies at the core of behavioral economics, without viewing descriptive *or normative* failures of rationality axioms as leading to new rationalizations for paternalistic policies (Sugden, 2004).

This normative debate will, no doubt, continue. We only wish to add an observation relevant for interpreting our finding that economists' beliefs about PSA testing and the risks of prostate cancer typically violate the assumption of Bayesian rationality. When normative theory and observed behavior come into conflict, behavioral economics typically follows the research program laid out in Tversky and Kahneman (1986) by unequivocally attributing error to the agent responsible for the behavior. That is, however, not the only valid deduction one can take away from this conflict between normative theory and observed behavior. One can instead conclude that principles previously thought to have normative value are simply incomplete, or perhaps have a more limited range of applicability, than previously thought.

Tversky and Kahneman (1986) put forward an analogy equating behavioral anomalies and optical illusions. Behavioral anomalies are anomalies because they deviate from axiomatic normative decision theory. Optical illusions are illusions because perceived distances deviate from objectively measured distance. The implication is that the axiomatic foundation of normative decision theory is as solidly grounded as the measure of physical distance.

Thaler (1991, p. 138) writes, "It goes without saying that the existence of an optical illusion that causes us to see one of two equal lines as longer than the other should not reduce the value we place on accurate measurement. On the contrary, illusions demonstrate the need for rulers!" Yet, in documenting (again and again) that observed behavior deviates from the assumptions (and predictions) of expected utility theory, there is no analog to the straight lines of objectively equal length. Unlike the simple geometric verification of equal lengths against which incorrect perceptions may be verified, the fact that human decisions do not satisfy the axioms underlying expected utility theory in no way implies an illusion or a mistake. Expected utility theory is, after all, but one model of how to rank risky alternatives. We would make the modest suggestion that behavioral economics could benefit from boldly pursuing new normative criteria that more effectively classify different procedures for making decisions in a way that helps assess whether they are well-matched to the environment in which they are used, according to the principle of ecological rationality (Gigerenzer and Selten, 2001; Smith, 2003).

References

Andreoni, J. and J. Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," Econometrica, 70 (2002), 737-53.

Andriole G.L., E.D. Crawford, R.L. Grubb, S.S. Buys, D. Chia, and T.R. Church TR, et al."Mortality

Results From A Randomized Prostate-Cancer Screening Trial.” N Engl J Med. 360(13) (2009):1310-9.
Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler, “Fact-Free Learning,” American Economic Review, 95 (2005), 1355-1368.
Ariely, D., G. Loewenstein, and D. Prelec, “Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences,” Quarterly Journal of Economics, 118 (2003), 73-106.
Baker, D.W., R.M. Parker, M.V. Williams, and W.S. Clark, "Health Literacy and the Risk of Hospital Admission," The Journal of General Internal Medicine, 13 (1998), 791-798.
Bardsley, N., R. Cubitt, G. Loomes, P. Moffatt, C. Starmer, and R. Sugden, Experimental Economics: Rethinking The Rules (Princeton: Princeton University Press, 2010).
Barry, Michael J, "The PSA Conundrum," Archives of Internal Medicine, 166 (2006), 38-43.
Baucells, M., J.A. Carrasco, and R.M. Hogarth, “Cumulative Dominance And Heuristic Performance In Binary Multi-Attribute Choice,” Operations Research, 56 (5) (2008), 1289-1304.
Behrens, Johann, Werner Güth, Hartmut Kliemt, and Vittoria Levati, "Games that Doctors Play Two-layered Agency Problems in a Medical System," in A. Mehler and U. Cantner (eds.), Dimensions of Public Economics: Strategies and Policies Facing Uncertainty and Market Failure (forthcoming).
Berg, Nathan, “Normative Behavioral Economics,” Journal of Socio-Economics, 32 (2003), 411-427.
Berg, N., C. Eckel, and C. Johnson, “Inconsistency Pays?: Time-Inconsistent Subjects And EU Violators Earn More,” Working Paper (2010), University of Texas-Dallas.
Berg, Nathan, and Ulrich Hoffrage, “Rational Ignoring with Unbounded Cognitive Capacity,” Journal of Economics Psychology, 29(6) (2008), 792-809.
Berg, Nathan and Gerd Gigerenzer, “Psychology implies paternalism?: Bounded rationality may reduce the rationale to regulate risk-taking,” Social Choice and Welfare, 28(2) (2007), 337-359.
Berg, Nathan and Gerd Gigerenzer, “As-if behavioral economics: Neoclassical economics in disguise?,” History of Economic Ideas, 18(1) (2010), 113-166.
Berg, Nathan and Donald Lien, "Does Society Benefit from Investor Overconfidence in the Ability of Financial Market Experts?" Journal of Economic Behavior and Organization, 58 (2005), 95-116.
Bewley, Truman, “Knightian Decision Theory: Part 1,” Decisions in Economics and Finance, 25 (2002), 79-110.
Binmore, K., Rational Decisions, (Princeton: Princeton University Press, 2008).
Binmore, K., J. McCarthy, G., Ponti, L., Samuelson, and A. Shaked, “A Backward Induction Experiment,” Journal of Economic Theory, 104 (2002), 48–88.
Blendon, Robert J., John M. Benson, Mollyann Brodie, Richard Morin, Drew E. Altman, Daniel Gitterman, Mario Brossard, and Matt James, "Bridging the Gap between the Public's and Economists' Views of the Economy," Journal of Economic Perspectives, 11(3) (1997), 105-118.
Boyd, Robert and Peter J. Richerson, Culture and the Evolutionary Process, (Chicago: University of Chicago Press, 1985).
Braga, J., S.J. Humphrey, and C. Starmer, “Market Experience Eliminates Some Anomalies-And Creates New Ones,” European Economic Review, 53 (2009), 401-416.
Camerer, Colin F. and Robin M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," Journal of Risk and Uncertainty, 19 (1999), 7-42.
Camerer, C., S. Issacharoff, G. Loewenstein, T. O’Donoghue, and M. Rabin, “Regulation For Conservatives: Behavioral Economics And The Case For “Asymmetric Paternalism.”” University of Pennsylvania Law Review, 1151 (2003), 1211-1254.
Camerer, C., G. Loewenstein, and M. Weber, “The Curse Of Knowledge In Economic Settings: An Experimental Analysis,” Journal of Political Economy, 97 (1989), 1232-1254.
Caplan, Bryan, "What Makes People Think Like Economists? Evidence on Economic Cognition from

the ‘Survey of Americans and Economists on the Economy,’" Journal of Law and Economics, 44(2) (2001), 395-426.
Caplan, Bryan, "Systematically Biased Beliefs About Economics: Robust Evidence of Judgmental Anomalies from the Survey of Americans and Economists on the Economy," The Economic Journal, 112(479) (2002), 433-458.
Carter, John R. and Michael D. Irons, "Are Economists Different, and If So, Why?," Journal of Economic Perspectives, 5(2) (1991), 171-177.
Cherry, T., T. Crocker, and J. Shogren, "Rationality Spillovers," Journal of Environmental Economics and Management, 45 (2003), 63-85.
Chu, Yun-Peng and Ruey-Ling Chu, "The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings," American Economic Review, 80 (1990), 902-911.
Ciatto, S., M. Zappa, R. Bonardi, and G. Gervasi, "Prostate Cancer Screening: The Problem of Overdiagnosis and Lessons to Be Learned from Breast Cancer Screening," European Journal of Cancer, 36 (2000), 1347-1350.
Ciatto, Stefano, "Reliability of PSA Testing Remains Unclear," British Medical Journal, 327 (2003), 750.
Comte, O. and A. Postlewaite, “Repeated Relationships With Limits On Information Processing,” PIER Working Paper 08-026 (2008), University of Pennsylvania.
Concato, John, Carolyn K. Wells, Ralph I. Horwitz, David Penson, Graeme Fincke, Dan R. Berlowitz, Gregory Froehlich, Dawna Blake, Martyn A. Vickers, Gerald A. Gehr, Nabil H. Raheb, Gail Sullivan, and Peter Peduzzi, "The Effectiveness of Screening for Prostate Cancer: A Nested Case-Control Study," Archives of Internal Medicine, 166 (2006), 38-43.
Cubitt, R. and R. Sugden, "Dynamic Decision-Making Under Uncertainty: An Experimental Investigation of Choices Between Accumulator Gambles," Journal of Risk and Uncertainty, 22(1) (2001), 103-128.
Davidson, D., J.C.C. McKinsey, and P. Suppes, "Outlines of a Formal Theory of Value, I," Philosophy of Science, 22 (1955), 140-160.
Dekel E. and S. Scotchmer, “On the Evolution of Attitudes Toward Risk in Winner-Take-All Games,” Journal of Economic Theory, 87 (1999), 95-124.
De Long, Bradford J., Andrei Shleifer, Lawrence H. Summers, and Robert Waldman, "The Survival of Noise Traders in Financial Markets," Journal of Business, 64 (1991), 1-19.
DeMiguel, V., L. Garlappi, and R. Uppal, “Optimal Versus Naive Diversification: How Inefficient Is The 1/N Portfolio Strategy,” Review of Financial Studies, 22 (2009), 1915–1953.
Draisma, Gerrit, Rob Boer, Suzie J. Otto, Ingrid W. van der Cruijsen, Ronald A.M. Damhuis, Fritz H. Schröder, and Harry J. de Koning, "Lead Times and Overdetection Due to Prostate-Specific Antigen Screening: Estimates from the European Randomized Study of Screening for Prostate Cancer," Journal of the National Cancer Institute, 95(12) (2003), 868-878.
Emerson, Ralph W., Selected Writings of Ralph Waldo Emerson, (New York, NY: Penguin, [1841] 2003).
Fehr, Ernst and Jean-Robert Tyran, "Individual Irrationality and Aggregate Outcomes," Journal of Economic Perspectives, 19 (2005), 43-66.
Frank, Robert H., Thomas D. Gilovich, and Dennis T. Regan, "Does Studying Economics Inhibit Cooperation?," Journal of Economic Perspectives, 7(2) (1993), 159-171.
Fuchs, Victor R., Alan B. Krueger, and James M. Poterba, "Economists' Views About Parameters, Values, and Policies: Survey Results in Labor and Public Economics," Journal of Economic Literature, 36(3) (1998), 1387-1425.
Gaechter, S., H. Orzen, E., Renner, and C. Starmer, C., “Are Experimental Economists Prone To Framing Effects? A Natural Field Experiment,” Journal of Economic Behavior and Organization, 70

(2009), 443-446.
Ganguly, Ananda, John H. Kagel, and Donald V. Moser, "Do Asset Market Prices Reflect Traders' Judgment Biases?," Journal of Risk and Uncertainty, 20(3) (2000), 219-245.
Gann, Peter H., "Interpreting Recent Trends in Prostate Cancer Incidence and Mortality," Epidemiology, 8 (1997), 117-120.
Gann, Peter H., C.H. Hennekens, and M.J. Stampfer, "A Prospective Evaluation of Plasma Prostate-Specific Antigen for Detection of Prostate Cancer," Journal of the American Medical Association, 273 (1995), 289-294.
Gigerenzer, Gerd, “Fast and Frugal Heuristics: The Tools of Bounded Rationality,” in D.J. Koehler and N. Harvey (eds.), Blackwell Handbook of Judgment and Decision Making. (Oxford: Blackwell, 2004, pp. 62-88).
Gigerenzer, Gerd and Ulrich Hoffrage, "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats," Psychological Review, 102(4) (1995), 684-704.
Gigerenzer, Gerd and Ulrich Hoffrage, "Overcoming Difficulties in Bayesian Reasoning: A Reply to Lewis, Meilers & McGraw," Psychological Review, 106 (1999), 425-430.
Gigerenzer, Gerd and Reinhard Selten, "Rethinking Rationality," in G. Gigerenzer and R. Selten (eds.), Bounded Rationality: The Adaptive Toolbox (Cambridge, MA: MIT Press, 2001, pp. 1-12).
Gilboa, Itzhak, “Questions In Decision Theory,” Annual Reviews in Economics, (forthcoming).
Gilboa, I., A. Postlewaite, and D. Schmeidler, “Probabilities n Economic Modeling,” Journal of Economic Perspectives, 22 (2008), 173-188.
Gilboa, I., A. Postlewaite, and D. Schmeidler, “Is It Always Rational To Satisfy Savage’s Axioms?” Economics and Philosophy, 25 (2009), 285-296.
Gilboa, I. and L. Samuelson, “Subjectivity In Inductive Inference,” Working Paper (2010), HEC-Paris.
Gilboa, I.,L. Samuelson, and D. Schmeidler, “Dynamics Of Inductive Inference In A Unified Model,” Working Paper (2010), Yale University.
Gilboa, Itzhak and David Schmeidler, “Case-Based Decision Theory,” Quarterly Journal of Economics, 110 (1995), 605-639.
Gintis, Herbert, “The Dynamics of General Equilibrium,” Economic Journal, 117 (2007), 1280–1309.
Gintis, Herbert, “Bayesian Rationality, Social Epistemology, And The Choreographer,” Working Paper (2010), Central European University.
Gintis, Herbert, “Rationality And Its Discontents,” Economic Journal (forthcoming).
Goldstein, D. G. and G. Gigerenzer, “Fast And Frugal Forecasting,” International Journal of Forecasting, 25 (2009), 760-772.
Grether, D.M. and Charles R. Plott, "Economic Theory of Choice and the Preference Reversal Phenomenon," American Economic Review, 69(4) (1979), 623-638.
Grunwald, P.D. and J.Y. Halpern, “When Ignorance Is Bliss,” ACM International Conference Proceeding Series, 70 (2004), 226-234.
Halpern, J.Y., “On Ambiguities In The Interpretation Of Game Trees,” Games and Economic Behavior, 20 (1997), 66-96.
Hammond, K.R., Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice, (New York: Oxford University Press, 1996).
Hammond, P., “Consequentialist Foundations For Expected Utility,” Theory and Decision, 25 (1988), 25–78.
Hastie, R. and K.A. Rasinski, “The Concept Of Accuracy In Social Judgment,” in D. Bar-Tal and A.W. Kruglanski (eds.), The Social Psychology of Knowledge (New York, NY; Paris, France: Cambridge University Press, 1988, pp. 193-208).
Haltiwanger, J.C. and M. Waldman, "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," American Economic Review, 75(3) (1985), 326-340.

Harris, Russell and Kathleen N. Lohr, "Screening for Prostate Cancer: An Update of the Evidence for the U.S. Preventative Taskforce Services," Annals of Internal Medicine, 37(11) (2002), 917-929.
Harbaugh, W.T., K. Krause, and T.R. Berry, "GARP for Kids: On the Development of Rational Choice Behavior," American Economic Review, 91 (2001), 1539-1545.
Hawkes, Nigel, "Prostate Cancer Test May Leave Men Even Worse Off," http://www.timesonline.co.uk/tol/news/uk/article786762.ece, (2006).
Heifetz, A. and E. Segev, "The Evolutionary Role of Toughness in Bargaining," Games and Economic Behavior, 49 (2004), 117-134.
Heifetz, A. and Y. Spiegel, "The Evolution of Biased Perceptions," Tel-Aviv University, (2001).
Hogarth, R. M., and N. Karelaia, "Simple Models For Multi-Attribute Choice With Many Alternatives: When It Does And Does Not Pay To Face Tradeoffs With Binary Attributes," Management Science, 51(12) (2005), 1860-1872.
Hogarth, R. M., and N. Karelaia, "'Take-The-Best' And Other Simple Strategies: Why And When They Work 'Well' With Binary Cues," Theory and Decision, 61 (2006), 205-249.
Holmberg, Lars, Anna Bill-Axelson, Fred Helgesen, Jaakko O. Salo, Per Folmerz, Michael Haggman, Swen-Olof Andersson, Anders Spangberg, Christer Busch, Steg Nordling, Juni Palmgren, Hans-Olov Adami, Jan-Erik Johansson, and Bo Johan Norlén, "A Randomized Trial Comparing Radical Prostatectomy with Watchful Waiting in Early Prostate Cancer," The New England Journal of Medicine, 347(11) (2002), 781-789.
Kitcher, Phillip, "The Naturalists Return," Philosophical Review,101 (1992), 53-114.
Kluger, Brian D. and Steve Wyatt, "Are Judgment Errors Reflected in Market Prices and Allocations? Experimental Evidence Based on a Monty Hall Problem," Journal of Finance, 59(3) (2003), 969-997.
Kyle, A.S. and A.Wang, "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?," Journal of Finance, LII (1997), 2073-2090.
Jolls, C., C.R. Sunstein, and R.H. Thaler, "A Behavioral Approach To Law And Economics." Stanford Law Review, 50 (1998), 1471-1541.
Laibson, D. and L.Yariv, "Safety in Markets: An Impossibility Theorem for Dutch Books," Harvard University, 2007.
Leland, J. "Generalized Similarity Judgments: An Alternative Explanation For Choice Anomalies," Journal of Risk and Uncertainty, 9 (1994), 151-172.
List, John and Daniel Millimet, "Bounding the Impact of Market Experience on Rationality: Evidence from a Field Experiment with Imperfect Compliance," Southern Methodist University, 2004.
Litwin, Mark S., Ron D. Hays, Arlene Fink, Patricia A. Ganz, Barbara Leake, Gary E. Leach, and Robert H. Brook, "Quality-of-Life Outcomes in Men Treated for Localized Prostate Cancer," The Journal of the American Medical Association, 273 (1995), 129-135.
Loewenstein, George, "Anticipation And The Valuation Of Delayed Consumption," The Economic Journal, 97(387) (1987), 666-684.
Loewenstein, George, "Experimental Economics from the Vantage-Point of Behavioural Economics," Economic Journal, 109(453) (1999), F23-34.
Loewenstein, George, "Hot-Cold Empathy Gaps and Medical Decision Making," Health Psychology, 24 (2005), S49-S56.
Loewenstein, George, "The Pleasures and Pains of Information," Science, 312 (2006), 704-706.
Loomes, G., C. Starmer, and R. Sugden, "Preference Reversal: Information Processing Or Rational Non-Transitive Choice?," Economic Journal, 99 (1989), 140-151.
Loomes, G., C. Starmer, and R. Sugden, "Observing Violations Of Transitivity By Experimental Methods," Econometrica, 59 (1991), 425-439.
Loomes, G., C. Starmer, and R. Sugden, "Preference Reversals And Disparities Between Willingness

To Pay And Willingness To Accept In Repeated Markets," Journal of Economic Psychology, 31 (2010)., 374-387.
Lusardi, A. and O.S. Mitchell, "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy And Retirement Readiness," Working Paper (2009), Dartmouth University.
McFall, Stephanie L. and Robert M. Hamm, "Interpretation of Prostate Cancer Screening Events and Outcomes: A Focus Group Study," Patient Education and Counseling, 49 (2003), 207-218.
Parker, Ruth M., David W. Baker, Mark V. Williams, and Joanne R. Nurss, "The Test of Functional Health Literacy in Adults: A New Instrument for Measuring Patients' Literacy Skills," Journal of General Internal Medicine, 10 (1995), 537-541.
Raiffa, H., Decision Analysis: Introductory Lectures on Choice under Uncertainty, (Reading, MA: Addison-Wesley, 1968).
Robson, A., and L. Samuelson, L., "The Evolution Of Time Preference With Aggregate Uncertainty," American Economic Review, 99 (2009), 1925-1953.
Rubinstein, Ariel, "Comments on the Interpretation of Game Theory," Econometrica, 59 (1991), 909-924.
Rubinstein, Ariel, "A Theorist's View of Experiments," European Economic Review, 45 (2001), 615-628.
Rubinstein, Ariel, ""Economics and Psychology"? The Case of Hyperbolic Discounting," International Economic Review, 44 (2003), 1207-1216.
Rubinstein, Ariel, "Dilemmas of An Economic Theorist. Econometrica," Econometrica, 74 (2006), 865-883.
Rubinstein, A. and M. Osborne, "Games with Procedurally Rational Players,"American Economic Review, 88 (1998), 834-847.
Rubinstein, Ariel and Ran Spiegler, "Money Pumps in the Market," Journal of the European Economic Association, 6 (2008), 237-253.
Samuelson, L., "Analogies, Adaptation, And Anomalies," Journal of Economic Theory, 97 (2001), 320–367.
Samuelson, L., and J. Swinkels, "Information, Evolution and Utility," Theoretical Economics, 1 (2006), 119-142.
Savage, L., The Foundations of Statistics, (New York, NY: Wiley, 1954).
Segal, U. and J. Sobel, J., "Tit For Tat: Foundations Of Preferences For Reciprocity In Strategic Settings," Journal of Economic Theory, 136 (2007), 197-216.
Selten, R., "What Is Bounded Rationality?," in G. Gigerenzer and R. Selten (eds.) Bounded Rationality: The Adaptive Toolbox (Cambridge, MA: MIT Press, 2001, pp. 13-36).
Sen, Amartya, "Internal Consistency of Choice," Econometrica, 61(3) (1993), 495-521.
Sippel, Reinhard, "An Experiment on the Pure Theory of Consumers' Behaviour," Economic Journal, 107 (1997), 1431-1444.
Smith, Vernon, "Constructivist and Ecological Rationality in Economics," American Economic Review, 93(3) (2003), 465-508.
Sorum, Paul Clay, Etienne Mullet, Junseop Shim, Sylvie Bonnin-Scaon, Gérard Chasseigne, and Joel Cogneau, "Avoidance of Anticipated Regret: The Ordering of Prostate-Specific Antigen Tests," Medical Decision Making, 24 (2004), 149-159.
Spiegel, Y., A. Heifetza, and C. Shannon, "What To Maximize If You Must," Journal of Economic Theory, 133 (2007), 31 – 57.
Stanford, J.L., R.A. Stephenson, L.M. Coyle, J. Cerhan, R. Correa, J.W. Eley, F. Gilliland, B. Hankey, L.N. Kolonel, C. Kosary, R. Ross, R. Severson, and D.West, "Prostate Cancer Trends 1973-1995," (Bethesda, MD: National Cancer Institute SEER Program, NIH Pub. No. 99-4543, 1999).
Starmer, Chris, "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of

Choice under Risk," Journal of Economic Literature, 38 (2000), 332-382.
Starmer, Chris., “Normative Notions In Descriptive Dialogues, Journal of Economic Methodology, 12 (2005), 277-290.
Starmer, Chris, "Friedman's Risky Methodology," in U. Mäki (ed.) The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy (Cambridge: Cambridge University Press, 2009, pp. 285-302).
Steineck, Gunnar, Fred Helgesen, Jan Adolfsson, Paul W. Dickman, Jan-Erik Johansson, Bo Johan Norlén, and Lars Holmberg, "Quality of Life after Radical Prostatectomy or Watchful Waiting," The New England Journal of Medicine, 347(11) (2002), 790-796.
Studdert, D. M., M.M. Mello, W.M. Sage, C.M. DesRoches, J. Peugh, K. Zapert, and T.A. Brennan, “Defensive Medicine Among High-Risk Specialist Physicians In A Volatile Malpractice Environment.” Journal of the American Medical Association, 293 (2005), 2609-17.
Sugden, R., “Rational Choice: A Survey Of Contributions From Economics And Philosophy,” Economic Journal, 101 (1991), 751–785.
Sugden, R., “The Opportunity Criterion: Consumer Sovereignty Without The Assumption Of Coherent Preferences." American Economic Review, 94 (2004), 1014–1033.
Sugden, R., “The Changing Relationship Between Theory And Experiment In Economics,” Philosophy of Science, 75 (2008a), 621–632.
Sugden, R., “Why Incoherent Preferences Do Not Justify Paternalism,” Constitutional Political Economy, 19 (2008b), 226–248.
Sugden, R. “Credible Worlds, Capacities and Mechanisms,” Erkenntnis, 70 (2009), 3–27.
Thaler, R.H., Quasi Rational Economics, (New York: Russell Sage Foundation, 1991).
Tukey, John, Exploratory Data Analysis, (Reading, MA: Addison-Wesley, 1977).
Tversky, Amos, "Intransitivity of Preferences," Psychological Review, 76 (1969), 31-48.
Tversky Amos and Kahneman Daniel, “Judgment Under Uncertainty: Heuristics And Biases,” Science, 185 (1974), 1124-1131.
Tversky Amos and Daniel Kahneman, “Rational Choice and the Framing of Decisions,” Journal of Business, 59 (1986), S251-S278.
U.S. Preventive Services Task Force, "Screening for Prostate Cancer: Recommendation and Rationale," Annals of Internal Medicine, 137(11) (2002), 915-916.
Wang, Linda and Katherine Arnold, "Prostate Cancer Incidence Trends Reveal Extent of Screening-Related Overdiagnosis," Journal of the National Cancer Institute, 94(13) (2002), 957.
Warneryd, K., “Sexual Reproduction And Time-Inconsistent Preferences,” Economics Letters (forthcoming).
Williams, Mark V., Ruth M. Parker, David W. Baker, Nina S. Parikh, Kathryn Pitkin, Wendy C. Coates, and Joanne R. Nurss, "Inadequate Functional Health Literacy among Patients at Two Public Hospitals," The Journal of the American Medical Association, 274(21) (1995), 1677-1682.
Wisdom, J., J.S. Downs, and G. Loewenstein, “Promoting Healthy Choices: Information Versus Convenience,” American Economic Journal: Applied Economics, 2 (2010), 164–178.
Yao, Siu-Long and Grace Lu-Yao, "Understanding and Appreciating Overdiagnosis in the PSA Era," Journal of the National Cancer Institute, July, 94(13) (2002), 958-960.
Yezer, Anthony M., Robert S. Goldfarb, and Paul J. Poppen, "Does Studying Economics Discourage Cooperation?: Watch What We Do, Not What We Say or How We Play," Journal of Economic Perspectives, 10(1) (1996), 171-186.

Table 1: Survey responses

individual characteristics	Fraction Yes	Number of Responses	
Keep $3 cash?	0.12	133	
Give $3 to charity?	0.71	133	
Chocolate?	0.17	133	
Economist?	0.92	133	
Work is applied as opposed to theoretical?	0.75	124	
Neoclassical methodological orientation?*	0.75	128	
50 years old or older**	0.62	133	
PSA decision and recommendation			
Did you have a PSA?	0.46	133	
Would you recommend a PSA to men in their 50s?	0.91	124	
information acquisition, perceived harms, and mode of information processing			
Written info?	0.22	131	
Medical journal?	0.05	131	
Harms?	0.25	122	
Weighed pros and cons?	0.36	128	
social influences			
Doctor influenced?	0.58	133	
Spouse or relative influenced?	0.07	133	
Nobody influenced?	0.15	133	

elicited probabilities	Mean Subjective Belief	Std Dev of Mean	Number of Responses	Published point-estimates***
lifetime incidence Pr(C Lifetime)	0.27	0.019	132	0.177
lifetime mortality Pr(D Lifetime)	0.06	0.006	132	0.028
posterior probability Pr(C|+)	0.47	0.019	128	0.34
sensitivity Pr(+|C)	0.72	0.018	126	0.68
incontinence probability Pr(Incontinence|Surgery)	0.30	0.020	128	.020 to 0.29

*Other individual information was collected too, for example, subfield specialization indicators used as controls in some regressions reported below. The sample of self-reported primary specializations consisted of 7 percent econometrics, 12 percent finance, 5 percent health economics, 7 percent economic history, 5 percent in industrial organization, and 9 percent macroeconomics. No subfield indicator correlates with neoclassical methodological orientation by more than 0.12, and some, like econometrics and economic history, have slight negative correlations with the neoclassical indicator. **All 133 respondents reported their age in years, 119 of whom were 40 or older. Mean self-reported age was 51 years old. ***Stanford et al's (1999) NCI SEER study and Harris and Lohr (2002).

Table 2: PSA controversy in the medical literature

Journal	Author(s)	Comment
Archive of Internal Medicine	Concato, et al (2006)	"Measurement of prostate-specific antigen (PSA) in serum and digital rectal examination (DRE) are commonly used to screen for prostate cancer, yet official recommendations regarding these tests vary. For example, American Cancer Society and American Urological Association recommendations include screening for prostate cancer in men older than 50 years, using PSA testing and DRE, followed by transrectal ultrasound if either test result is abnormal. In contrast, the American College of Physicians suggests counseling regarding possible benefits and risks, and the US Preventative Services Task Force found insufficient evidence to recommend screening. These positions were promulgated in the setting of data showing that the screening tests increase detection of prostate cancer but without direct evidence showing that PSA or DRE reduce mortality."
Annals of Internal Medicine	Barry (2006)	"We already know that PSA screening has a substantial downside. . . .The poor specificity of PSA testing results in a high probability of false positives requiring prostate biopsies and lingering uncertainty about prostate cancer risk, even with initially negative biopsy findings. Although we now know that aggressive surgical treatment of prostate cancers largely detected the "old fashioned way" without screening has a modest benefit, with about 18 cancers needing to be removed to prevent 1 death over 10 years, that benefit comes at a considerable price in terms of sexual dysfunction and incontinence. The key question is whether early detection and subsequent aggressive treatment of prostate cancers found through PSA screening prevents enough morbidity and mortality to overcome these disadvantages..."
Journal of the National Cancer	Draisma et al (2003)	"Whether asymptomatic men benefit from screening for prostate cancer is an unresolved question."
New England Journal of Medicine	Steineck et al (2002)	Regarding watchful waiting versus other treatment options following a diagnosis of prostate cancer, the "alternatives are associated with complex and incommensurable outcomes, and each man must judge for himself which treatment is preferable."
European Journal of Cancer	Ciatto et al (2000)	"The benefits of prostate cancer screening are just theoretical, thus far unknown, and the potential risk of adverse effects much more worrying than for breast cancer: screening as a current practice is unethical, and the practice of screening, at the moment, must be limited to experimental studies." [also see Ciatto (*British Medical Journal,* 2003)]
American College of Physicians	Concato (1999)	"Routine PSA measurement without a frank discussion of the issues involved is inappropriate."
Epidemiology	Gann (1997)	"The most important question is whether the decline in [disease-specific] mortality* will be worth the cost--in terms of anxiety, excess biopsies, and even unnecessary surgery." [also see Gann et al (*JAMA* , 1995)]
Journal of the American Medical Association (JAMA)	Litwin et al (1995)	Regarding patients' treatment decisions and doctors' recommendations: "Little is known about how or why they make treatment decisions, how their quality of life is affected by therapy, or why physicians recommend one treatment vs. another." Regarding costs and benefits: "The traditional Western medical perspective of maximizing survival at all cost is inadequate. Indeed, the most rational approach to treating men with localized prostate cancer needs to include not only adding years to life, but also adding life to years."

*The most common recommendation appears to be that doctors should provide patients with information about the PSA test's pros and cons and encourage patients to decide about PSA testing on an individual basis. Medical communication experts refer to this as the balance-sheet approach, advocated for the purpose of encouraging patients to weigh costs and benefits rather than making automatic decisions in favor of screening or treatment (Concato, 1999; McFall and Hamm, 2003). The National Cancer Institute (the cancer wing of the U.S. National Institutes of Health) explicitly recommends against routine screening of asymptomatic men, and its website (www.cancer.gov) states that men should consider costs and benefits before deciding on a PSA test. In contrast, many hospitals and doctors adopt a policy of automatic screening, as is recommended by the American Cancer Society and the American Urological Association. The recommendation to weigh costs and benefits before being tested is echoed in numerous medical journal articles, however, including most of those listed in Table 1.

Table 3: Contrasts in mean inaccuracy between consistent and inconsistent subsamples

		consistent	inconsistent		consistent	inconsistent		consistent	inconsistent		consistent	inconsistent	
	Grand Mean	24 Perfect Bayesians	101 Deviators from Bayes	t stat	60 strictly below median incon-sistency	65 weakly above median incon-sistency	t stat	36 weakly below 25th percentile incon-sistency	34 weakly above 75th percentile incon-sistency	t stat	80 Ballpark Bayesians	45 Emersonians	t stat
Measures of inaccuracy													
inaccuaracy	0.99	1.26	0.90	1.7	1.08	0.87	1.6	1.26	0.77	2.5	1.08	0.78	2.2
signed inaccuracy	0.01	-0.56	0.16	-2.2	-0.12	0.15	-1.3	-0.45	0.04	-1.6	-0.14	0.32	-2.1
Log deviations of individual elicited beliefs													
log(incidence/0.177)	-0.06	-0.43	0.08	-1.7	-0.13	0.09	-1.0	-0.44	-0.04	-1.3	-0.11	0.15	-1.2
log(mortality/0.028)	0.07	-0.69	0.23	-2.2	-0.11	0.21	-1.2	-0.48	0.11	-1.5	-0.18	0.48	-2.5
log(posterior/0.34)	0.18	0.00	0.22	-2.1	0.11	0.23	-1.1	0.09	0.12	-0.2	-0.11	0.67	-7.9
log(sensitivity/0.64)	0.06	0.06	0.07	0.0	0.11	0.02	1.5	0.13	-0.10	2.5	0.09	0.01	1.3
Measures of inconsistency													
inconsistency	0.48	0.00	0.59	--	0.12	0.81	--	0.03	1.05	--	0.34	0.73	--
signed inconsistency	-0.17	0.00	-0.21	--	-0.06	-0.28	--	-0.02	-0.28	--	0.14	-0.73	--

*Inaccuracy is the (within-individual) simple average of the four absolute log deviations. Signed inaccuracy is the simple average of those same log deviations without taking absolute values. **Inconsistency is the absolute log-approximated percentage error of the elicited ratio, *posterior/sensitivity,* relative to the correct ratio of 1/2. Signed inconsistency is the same as inconsistency but without absolute values.

Table 4: Regression of inaccuracy

predictors	coef	t
consult written?(1/0)	0.35	2.0
consult med j?(1/0)	-0.31	-0.9
procon?(1/0)	-0.40	-2.8
times guess 50 (2,1,0)	-0.10	-1.1
nobody influenced?(1/0)	0.10	0.5
doctor influenced?(1/0)	-0.07	-0.4
age	-0.08	-1.2
age squared	0.00	1.3
psa (1/0)	0.04	0.3
cash (1/0)	-0.13	-0.7
chocolate (1/0)	-0.02	-0.1
noneconomist (1/0)	0.33	1.1
neoclassical?(1/0)	-0.31	-2.0
applied?(1/0)	-0.02	-0.1
econometrics(1/0)	-0.36	-1.3
finance(1/0)	-0.01	0.0
health economics(1/0)	-0.26	-0.9
history(1/0)	-0.08	-0.3
industrial organization(1/0)	0.42	1.3
labor(1/0)	0.48	2.1
macroeconomics(1/0)	-0.08	-0.3
inconsistency	-0.08	-0.6
constant	3.11	2.0
R2	*0.24*	
Sample Size	*117*	

Table 5: Estimated linear probability models for the PSA decision and PSA recommendation

	Empirical Models of the PSA Decision:						PSA Recommendation	
	fundamental		*add info-*		*add influencers*			
predictors	coefficient	t	coefficient	t	coefficient	t	coefficient	t
log(incidence/0.177)	0.05	1.0	0.07	1.4	0.04	0.9	-0.11	-2.4
log(mortality/0.028)	-0.01	-0.3	0.00	0.1	0.01	0.3	0.10	2.8
log(posterior/0.34)	-0.09	-1.6	-0.06	-0.9	-0.05	-0.7	-0.05	-0.7
log(sensitivity/0.64)	0.10	1.0	0.14	1.2	0.16	1.5	0.18	1.4
log(incontinence/0.150)	-0.06	-1.6	-0.07	-1.7	-0.08	-2.3	-0.07	-2.7
age	-0.03	-1.1	0.00	0.1	-0.02	-0.6	0.02	0.7
age squared	0.00	2.0	0.00	0.6	0.00	1.3	0.00	-0.7
cash?(1/0)			-0.15	-1.5	-0.17	-2.0	-0.10	-0.9
chocolate?(1/0)			-0.08	-0.7	-0.09	-0.8	-0.08	-0.9
procon?(1/0)			-0.06	-0.6	-0.04	-0.4	-0.05	-0.6
consult written?(1/0)			0.14	1.5	0.15	1.6	0.13	1.4
inconsistency			0.01	0.2	0.00	0.1	-0.02	-0.3
nobody influenced?(1/0)					-0.09	-0.7	-0.17	-1.3
doctor influenced?(1/0)					0.27	2.9	-0.03	-0.3
constant	0.79	1.0	-0.09	-0.1	0.52	0.6	0.65	0.8
R2	*0.34*		*0.38*		*0.46*		*0.18*	
Pr(test stat>observed|H0)	*0.13*		*0.14*		*0.03*		*0.01*	
Sample Size	*121*		*114*		*114*		*114*	

*H0 is the joint hypothesis that the first five variables, which proxy for perceived costs and benefits, have zero effect on the probability of having (or recommending) a PSA. The test statistic is distributed as F(5, sample size minus number of regressors) under the null.

Table 6: Cross-tabulation of harms of PSA and weighing pros and cons

		Would you say you weighed pros and cons in making your decision about whether to have a PSA?			
		no	yes	no response	Total
In your opinion are there potential harms associated with PSA screening?	no	59 [25]*	30 [16]	3 [1]	92 [42]
	yes	16 [9]	12 [7]	2 [1]	30 [17]
	no response	7 [2]	4 [0]	0 [0]	11 [2]
	Total	82 [36]	46 [23]	5 [2]	133 [61]

*Bracketed counts refer to the number of respondents in each cell who reported having had a PSA.

Table 7: PSA decisions broken out by information sources and self-reported information processing

Among those who:	Took PSA and 50+?*** yes	no	fraction yes
report that nobody influences them	4	7	0.36
report that someone influences them	43	12	0.78
somebody influences & NOT weigh pros and cons	26	8	0.76
somebody influences & weigh pros and cons	16	4	0.80
do NOT weigh pros and cons*	31	20	0.61
weigh pros and cons*	22	7	0.76
weigh pros and cons & report NO harms	16	21	0.76
weigh pros and cons & report harms	6	7	0.86
do NOT weigh pros and cons & report harms	7	4	0.64
weigh pros and cons & report harms	6	1	0.86
do not consult written sources**	34	28	0.55
consult written sources**	19	1	0.95
consult written sources & NOT weigh pros and cons	10	1	0.91
consult written sources & weigh pros and cons	9	0	1.00
perfect Bayesians	10	4	0.71
Emersonian (severe violations of probability theory)	18	11	0.62

*Among respondents age 50 and over, there were three who would not say whether they weighed pros and cons or not. Among these three, one reported having taken a PSA and two reported having taken no PSA. Under the heading "Took PSA and 50+?," those three non-responders (on the pros-versus-cons sample item) explain why the sums of "no"s and "yes"s across the "do NOT weigh pros and cons" and "weigh pros and cons" rows do not quite equal that of the row labeled "Among all." A similar explanation applies to the columns under the heading "Recommend PSA?" **Among the 29 respondents who said they consulted written information, 14 said they did not weigh pros and cons, and 15 said they did. ***The overall rate of PSA taking among respondents 50 and older was 65

Figure 1: Given the unconditional probability of a positive PSA test, P(+), and the unconditional probability of prostate cancer, P(C), two conditional beliefs are elicited: the posterior probability $P(C|+)_i$ and sensitivity $P(+|C)_i$

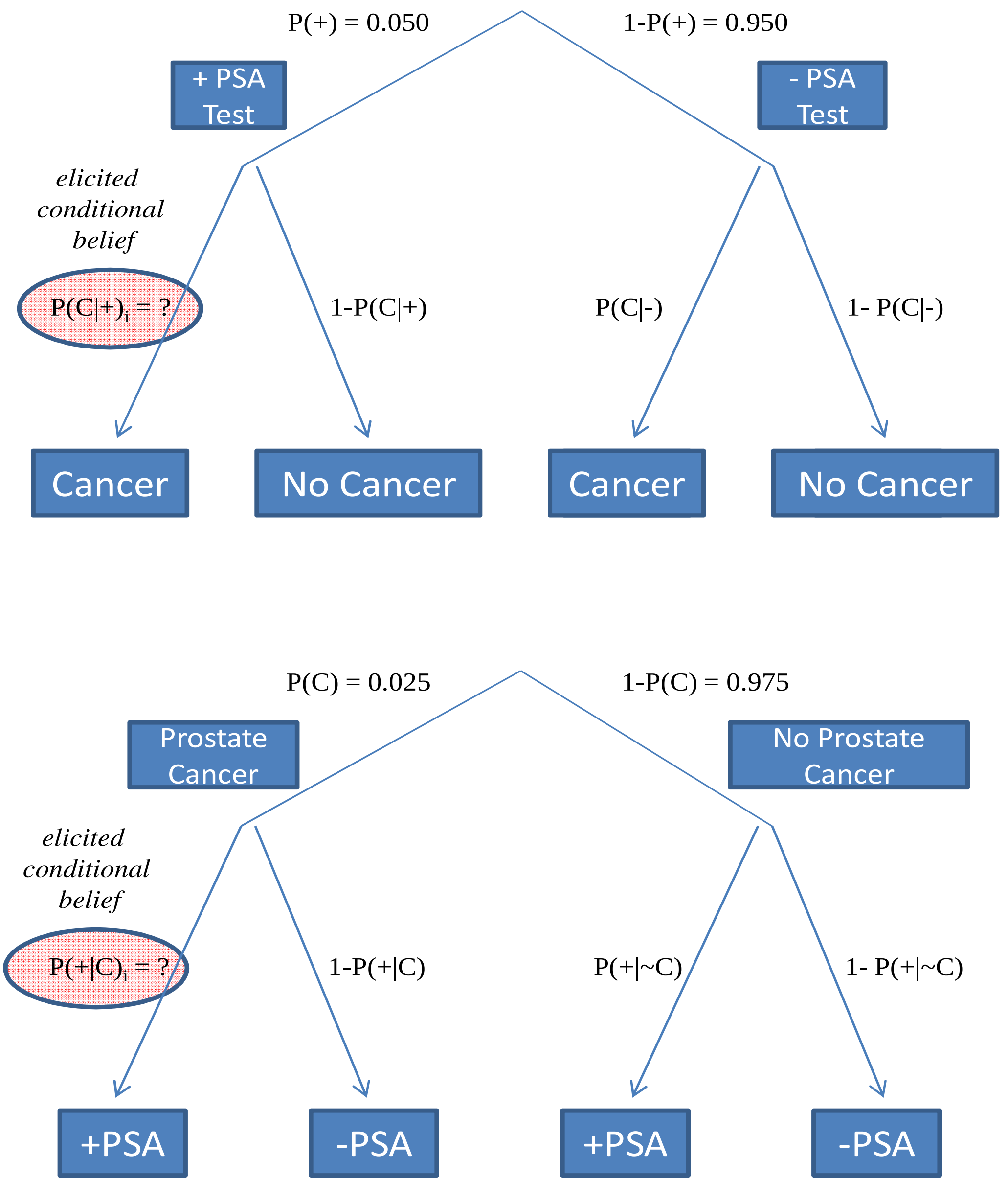
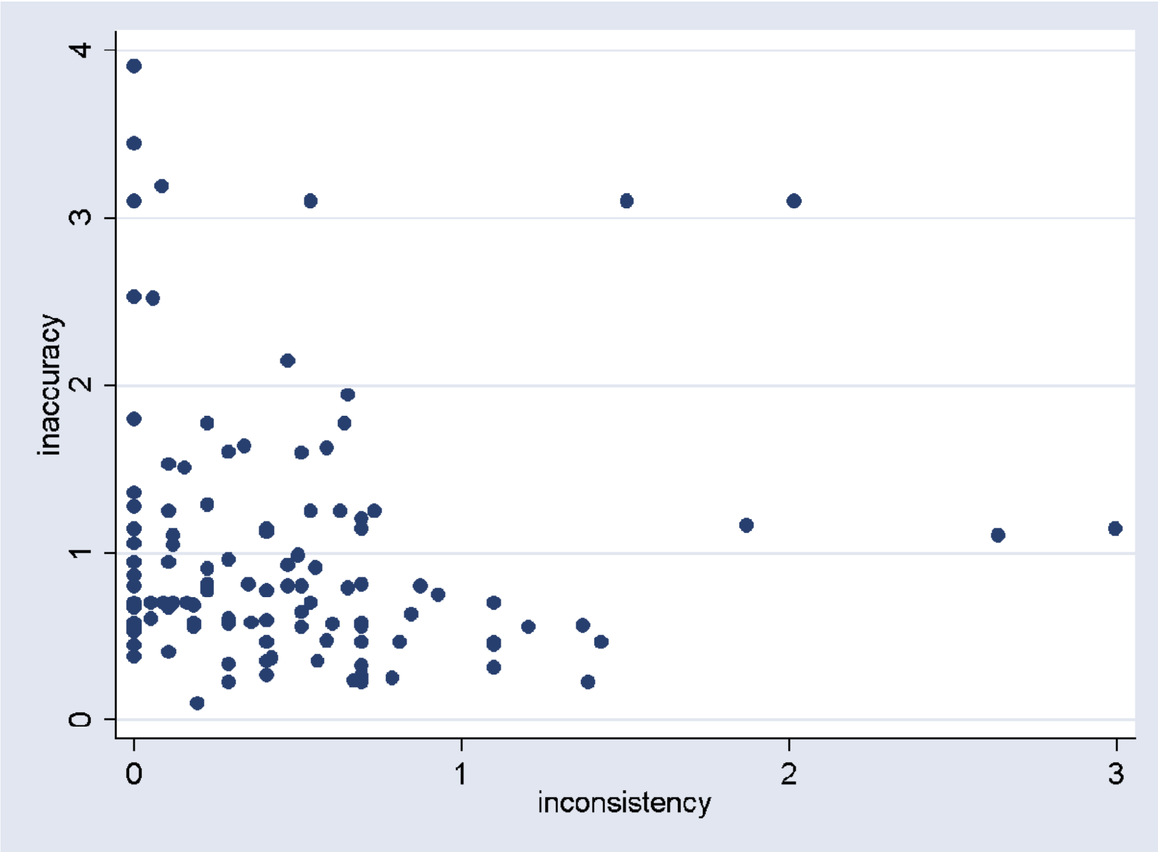


Figure 2: Inconsistency versus Inaccuracy (N=125)



*Bivariate regression line: *inaccuracy* = 1.00 - 0.06**inconsistency*. Because inconsistency and inaccuracy are defined as log deviations, the coefficient -0.06 can be interpreted as the elasticity of absolute inaccuracy (percentage-point deviation from published incidence and mortality rates) with respect to inconsistency (absolute percentage-point deviation from Bayes Rule). Simple correlation is -0.042.

NOT FOR PUBLICATION

Appendix 1: Survey Instrument

1. I'm conducting a survey about health decisions among economists and the first question is whether you'd like $3 or a Swiss chocolate.	*Three dollars for charity* □ *Chocolate* □ *Three dollars for self* □
2. Are you an economist?	***Yes***□ ***No***□
3. What's your subfield (within economics)?	
4. Would you say your professional work in economics is more theoretical or applied?	***Theoretical***□ ***Applied***□
The main focus of the survey is prostate cancer and PSA (Prostate Specific Antigen) screening. I won't ask any personal questions about the illness itself, just about screening. I'd like to elicit your best guesses about the risks of prostate cancer.	
5. For a randomly drawn American male, I'd like you to guess the probability that he will *be diagnosed* with prostate cancer in his lifetime?	
6. What would you say is the probability that he will die from prostate cancer in his lifetime?	
Now I'm going to ask you about American males in their 50s who have no symptoms, have never been diagnosed with prostate cancer, and are screened with a PSA test for the very first time. One leading study suggests that 5% of randomly sampled men from this population have a positive PSA. It's also estimated that 2.5% actually have prostate cancer at the time of screening, which includes those whose PSAs failed to detect the disease. [source: Harris et al, 2002, Ann Intern Med]	
7. Given a positive PSA, I'd like you to estimate the probability that a man actually has prostate cancer.	
8. And given cancer at the time of screening, what would you say is the probability of a positive PSA?	
9. In your opinion are there potential harms associated with PSA screening? If so, what are they?	***Yes***□ ***No***□ Potential harms include:
10. Now I'd like you to consider a man in his 50s whose PSA test detected prostate cancer *and* who was treated with surgery. What would you guess is the probability that he will suffer from incontinence as a result of the treatment?	
11. Did you ever have a PSA screening for prostate cancer? If yes, how many times?	**Yes**□ **No**□ # times ________
12. Whose views contributed to your decision about whether to have the PSA screening?	
13. Did you consult any written sources of information in making your decision?	
14. Did you consult any authoritative medical sources such as medical journals? If so, which source(s)?	**Yes**□ **No**□ Sources:
15. Would you say you weighed pros and cons in making your decision about whether to have a PSA?	***Yes***□ ***No***□
16. Would you recommend that men in their 50s take a PSA?	***Yes***□ ***No***□
17. How old are you?	**______ years old, or:** □ **age<40**, □**40-49,** □**50-59,** □**60-69, 70+**□
18. Would you consider yourself a neoclassical economist?	***Yes***□ ***No***□

NOT FOR PUBLICATION

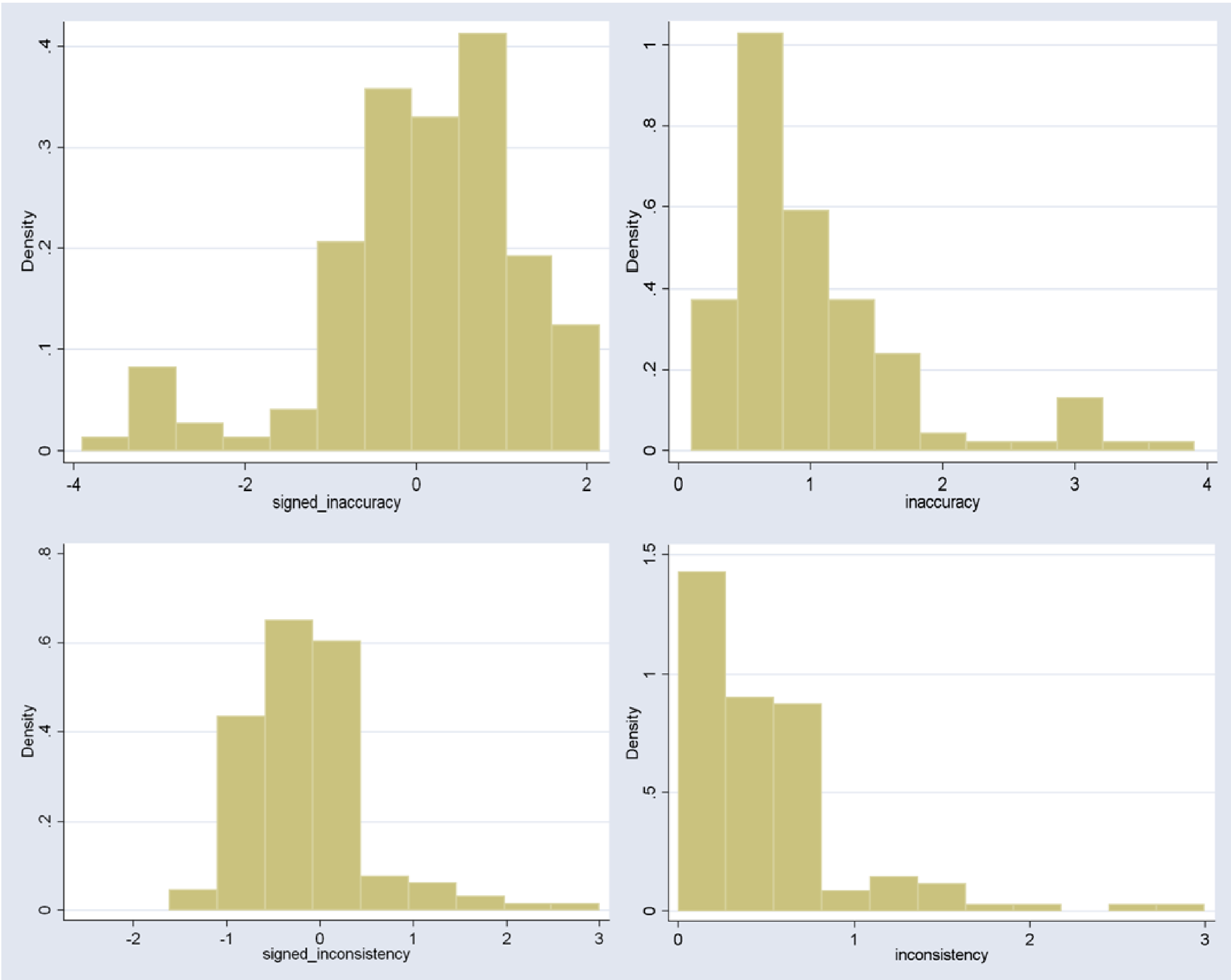
Appendix 2: Signed Versus Absolute Measures of Inconsistency and Inaccuracy, and Interpretation of Units in the Log Approximation

So far, deviations have been computed using absolute value. Signed versions of inconsistency and inaccuracy were constructed to see if useful information might be contained in the signs of deviations. We constructed an analogous pair of variables, *signed inconsistency* and *signed inaccuracy*, with identical definitions to *inconsistency* and *inaccuracy*, but without absolute values. The Figure in Appendix 3 shows empirical distributions for *signed inaccuracy*, *inaccuracy*, *signed inconsistency* and *inconsistency*.

A few examples help interpret the units of *inaccuracy* and *inconsistency.* An individual with *inaccuracy* = 0.10 provided mortality and incidence beliefs that were, on average, 10 percent too large *or* too small. The log approximations of percentage deviations become imprecise for large deviations. For example, the individual in our sample with *signed inaccuracy* =-3.9 reported mortality and incidence beliefs that were 2 percent of published point estimates, that is, 98 percent too small rather than "390 percent too small," since -3.9 = log(0.02). The exact percentage deviation, [(*incidence/0.177 -1)* + (*mortality/0.028-1)*]/2, has an empirical range of -0.98 to 9 (i.e., some respondents' beliefs are 98 percent too small while others have beliefs that are 900 percent too large).

NOT FOR PUBLICATION

Appendix 3: Empirical distributions of unsigned inaccuracy, unsigned inconsistency, inaccuracy and inconsistency



NOT FOR PUBLICATION

Appendix 4: Natural Frequencies

No Natural Frequency Effect

Previous studies have documented large differences in decisions resulting from logically equivalent representations of statistical information (e.g., the framing effect in Tversky and Kahneman, 1986). Gigerenzer and Hoffrage (1995, 1999) showed that communicating probabilities in natural frequencies (e.g., "7 in 1000" versus "0.7 percent") can lead to dramatic improvements in Bayesian reasoning. Our interview protocol alternated between two versions of the interview script which varied the way that probabilities were communicated to, and elicited from, respondents. In the probability treatment, respondents were told that "2.5 percent have prostate cancer," whereas in the natural frequency treatment they were told that "25 in 1000 have prostate cancer."

Counter to our expectations, the data showed virtually no treatment effect. In hindsight, we might have expected no effect because of a key difference between our elicitation and those for which large treatment effects have been shown previously. An important advantage of natural frequencies is that the reference class is held constant, making conditional probabilities easier to understand for those without statistical training (e.g., "50 in 1,000 had a positive PSA and 17 of those 50 actually had cancer" may be easier to understand than "the probability of a positive PSA is 0.05 and the probability of cancer conditional on a positive PSA is 0.34 [=17/50]"). Our elicitation scheme, however, switched between three different reference classes: 1,000 randomly drawn U.S. adult males (when eliciting *incidence* and *mortality*); 1,000 randomly drawn 50-year-olds without any symptoms or history of prostate cancer being screened for the first time (when eliciting *posterior probability* and *sensitivity*); and 1,000 randomly drawn

U.S. males who have been diagnosed with prostate cancer and treated with surgery (when eliciting the probability *incontinence*). Because the reference classes change, it is little surprise that natural frequencies did not improve Bayesian reasoning. On the other hand, it could be that economists' specialized training enabled them to interpret probabilistic and natural frequency representations more or less equivalently.