



BUSINESS SCHOOL
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1308

April 2013

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

Berg, N., Biele, G. and Gigerenzer, G.

Address for correspondence:

Nathan Berg
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: nathan.berg@otago.ac.nz
Telephone: 64 3 479 8648

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)

Abstract: When economists' subjective beliefs about the sensitivity and positive predictive value of the Prostate Specific Antigen (PSA) test are internally consistent (i.e., satisfying Bayes' Rule), their beliefs about prostate cancer risk are less accurate than among those with inconsistent beliefs. Using a loss function framework, we investigate but cannot find evidence that inconsistent beliefs lead to inaccuracy, different PSA decisions, or economic losses. Economists' PSA decisions appear to depend much more on the advice of doctors and family members than on beliefs about cancer risks and the pros/cons of PSA testing, which have little to no joint explanatory power.

Section 1: Introduction

Consistency of prior and posterior beliefs (i.e., conforming to Bayes' Rule) is a predominant normative characterization of rational beliefs.¹ 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) observes that before non-additive probability models appeared in the economics literature, economists usually took it for granted that the Savage Axioms (guaranteeing 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.” Selten (2001, p. 13) writes that “[m]odern 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, pp. 1214-1215) definition of “full rationality” requires that

¹ Savage argued for a normative interpretation of expected utility theory while admitting that he himself violated the theory when first encountering pairs of gambles in Allais' paradox (Savage, 1954). See Starmer (2000, 2009) for more on normative interpretations of expected utility theory.

“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 Aragonés et al., (2005, p. 1364), “[m]ost 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 of belief updating procedures), stating that “within economic theory the Bayesian approach is the sole claimant to the throne of rationality.”²

It is well known, however, that subjective beliefs can be logically consistent (i.e., perfectly Bayesian) and yet completely wrong (i.e., inaccurate). Surprisingly, there is, as yet, little empirical evidence of links between internal logical consistency of subjective beliefs and objective accuracy. This paper introduces a new elicitation technique that simultaneously measures deviations of an individual's beliefs from Bayes' Rule and deviations from objective frequencies. The belief data presented below provide some of the first empirical evidence regarding consistency versus accuracy of subjective beliefs.

If consistency and accuracy were positively correlated, then the literature documenting deviations from Bayesian beliefs would have new implications for decision making in environments that reward objective accuracy. To the extent that consistency and accuracy are unrelated, however, the normative force of Bayes' Rule in settings where accuracy rather than consistency is rewarded may be called into question.³

² Gintis (forthcoming, p. 2) states strong support for Bayesian consistency as a universal assumption: “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.’” Loewenstein (2006) urges caution, however, suggesting that theoretical extensions of standard models in pursuit of added realism, especially concerning information and information processing, do not necessarily wind up being more realistic. Cubitt and Sugden (2001) present novel experimental data in which inconsistent individuals do not necessarily succumb to exploitative competitors.

³ Psychologists Hastie and Rasinski (1986) appear to be the first to have classified normative criteria according to whether they require internal consistency (i.e., coherence) or correspondence to a free-standing performance measure such as objective accuracy. Hastie and Rasinski (1986) and Hammond (1996) refer to norms based on internal consistency as *coherence norms*, which include Bayes' Rule, the Kolmogorov axioms, and transitive preferences. In contrast, they categorize as *correspondence norms* those normative criteria that evaluate beliefs and decisions by how well they

The empirical evidence presented in this paper is intended to provide a first step toward investigating the relationship between these distinct normative criteria (i.e., consistency and accuracy) for evaluating the quality of subjective beliefs. This paper introduces and implements a technique for reporting empirical regularities (if any are present) linking consistency and accuracy. The elicitation technique allows for virtually any functional relationship between consistency and accuracy, thus facilitating a test of the hypothesis: Do people with consistent beliefs also tend to have accurate beliefs?

A second issue addressed in this paper concerns whether inconsistency is associated with economic losses. Despite the vast literature on non-Bayesian beliefs, one finds surprisingly little direct evidence to substantiate the hypothesis that deviations from Bayes' Rule generate meaningful economic losses.⁴ Raising questions about whether these deviations from standard normative benchmarks are costly (or perhaps even beneficial) should not imply broader skepticism about the 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?

To organize the data in a way that can potentially reveal economic losses due to inconsistent beliefs in the context of PSA testing, we adopt a model which assumes that decisions are based on minimization of a loss function. The loss function depends on beliefs about prostate cancer risks,

correspond to the demands of the decision maker's environment, bringing in a free-standing metric to measure the level of performance in different contexts (e.g., accuracy of beliefs, accumulated wealth, lifespan, or happiness). Unlike coherence norms which impose restrictions on pairs or larger sets of beliefs (and decisions) belonging to a single decision maker, correspondence norms enable interpersonal rankings on the basis of a single observation from each decision maker. Gilboa (forthcoming) poses the question of whether norms *not* based on internal consistency belong in definitions of economic rationality at all and makes a strong case for considering criteria other than consistency.

⁴ Behavioral economists have paid close attention to modeling and empirically documenting deviations from Bayes' Rule (e.g., Camerer 1987, 1992; Ganguly, Kagel and Moser, 2000; Kluger and Wyatt, 2003). The tacit motivation behind much of this literature is the normative assumption that people would be better off if their beliefs adhered more closely to Bayes' Rule.

beliefs about the quality of the screening instrument, and the PSA decision itself. We follow Gaechter, Orzen, Renner, and Starmer (2009) in taking advantage of economists' high levels of statistical fluency and familiarity with the axioms of rational choice by analyzing data collected from them.

Minimization of expected losses leads to an objective function that can be influenced by inconsistent beliefs through two distinct channels that are not mutually exclusive. The first potential channel would be inconsistency causing inaccurate beliefs (that lead to greater expected losses than is achieved by those with consistent beliefs). In this case, one expects to find a positive statistical association between inconsistency and inaccuracy.

The second theoretical channel for inconsistency to potentially cause losses would be if inconsistent individuals made substantially different choices conditional on the same belief profile. The second case would occur if, conditional on the same subjective beliefs, non-Bayesians tended to make different decisions (whether or not to take a PSA test) than Bayesians. Thus, the second potential channel would associate inconsistent beliefs with a distinct mapping from beliefs into action. In this case, one would expect to find an effect of inconsistency on the PSA decision itself (either positive or negative, since we make no assumption about whether PSA testing is beneficial or detrimental) after controlling for subjective beliefs about cancer risks and the benefits and costs of PSA testing.

Section 2 describes the data. Section 3 reports statistical models linking consistency to accuracy of beliefs and PSA decisions. Section 4 concludes with a discussion of interpretations of these results and the role of consistency in normative behavioral decision science.

Section 2: Description of data

Summary Statistics of 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. 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 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 that served as a passageway to the conference sessions.

The interviewer approached men who appeared at least 40 years old. An introductory statement offered respondents a choice of \$3 cash or a Swiss chocolate bar, together with 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 chose to do. Table 1 provides summary statistics of the survey responses used in subsequent statistical models.

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. Three quarters of respondents described their work as applied rather than theoretical. Three quarters of respondents also described their methodological approach as neoclassical (although Applied and Neoclassical were almost perfectly uncorrelated, with pairwise correlation of 0.0108). No respondent non-responded when asked to provide their age. The age distribution was remarkably symmetric covering a large range (26 to 79) with a mean of 51 and a strong majority (119 respondents) with ages 40 and above indicating that our interviewer largely succeeded at hitting the over-40 target.

Table 1 shows that roughly half the respondents (46 percent) reported having had a PSA. Among those 50 and older, the rate of PSA testing was 65 percent. When asked whether they recommend that asymptomatic men in the 50s take the PSA test as a screening device, most respondents (91 percent of the 124 who responded) responded affirmatively, with almost no difference in rates of recommendation by age. Summarized in the caption of Table 1 is information about

respondents' primary subfields of specialization.

Non-Response

The column labeled "Number of Valid Responses" shows that item non-response was a problem for several survey items although not the ones we would have expected. Nine men refused to classify their work as either "more applied" or "more theoretical." And nine refused to make a recommendation about whether men in their 50s should have a PSA. No one, however, refused to say whether he had taken a PSA.

Information acquisition, perceived harms and information processing

Table 1 shows that 22% of respondents reported having consulted written information . Only 5% reported having read a published article in a medical journal about PSA testing. The survey item labeled "Harms?" codes 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 and discussed below in Table 2), one might find it surprising that only a quarter of respondents said there were harms associated with PSA testing. More surprising, however, was that only about a third of respondents reported having weighed pros and cons when deciding whether to have a PSA test.

Not weighing pros and cons could, of course, be rationalized if someone perceived zero costs or zero benefits. In that case, there are no tradeoffs to consider. We worried, in fact, when designing the survey that the weighing pros and cons item might not generate any variation at all (with nearly all respondents answering "Yes"). Among the 30 respondents who said there were harms from PSA testing, 16 reported not weighing pros and cons. And among the 92 who said there were no harms (there were also 11 non-responses to the harms item not included in the rates reported here), 30 reported having weighed pros and cons.

Elicited Frequencies

The following five probabilistic beliefs were elicited which consist of two unconditional and three conditional probabilities:

- *lifetime incidence*: the probability that a randomly drawn male in the U.S. is diagnosed with prostate cancer within his lifetime, denoted $P(C \text{ Lifetime})$;
- *lifetime mortality*: the probability that a randomly drawn male in the U.S. dies of prostate cancer within his lifetime, denoted $P(D \text{ Lifetime})$;
- *incontinence probability*: the probability of incontinence conditional on surgical treatment for prostate cancer, denoted $P(\text{Incontinence} \mid \text{Surgery})$;
- *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 having undiagnosed prostate cancer at the time of screening, denoted $P(+|C)$.

The bottom block of elicited beliefs and published point estimates in Table 1 indicates that respondents' beliefs about these probabilities tended to be slightly too large, but not terribly far off from point estimates published in the *National Cancer Institute's Surveillance Epidemiology and End Results [SEER]* database (Stanford et al 1999) and *Annals of Internal Medicine* (Harris and Lohr 2002). Insofar as cost-benefit considerations are based on subjective beliefs about risk and the quality of the PSA test, 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.

Shifting Evidence and Expert Opinion That Make PSA an Important Decision to Study

Before introducing measures of consistency and accuracy of beliefs, Table 2 summarizes eight

frequently cited medical studies about the risks and benefits of PSA testing with quotations highlighting new statistical findings and shifting expert opinion (e.g., emerging from the National Institutes of Health (NIH)) that now recommends against automatic screening of asymptomatic men. Instead of automatic screening based on reaching an age threshold, the NIH recommendation is for men to weigh the benefits and costs of having a PSA test and make the decision on an individual basis in consultation with his doctor. The purpose of Table 2 is not to take a particular normative stand regarding PSA testing. Rather, Table 2 motivates interest in PSA testing as a potentially high-stakes decision task of interest to the decision science literature because of the divergence between the daily practice of many doctors in the US (e.g., automatically testing all men once they reach the age of 40) and the NIH recommendation of weighing pros and cons individually.

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. By the late 1990s, as many as half of American men over the age of 50 were estimated to have undergone PSA testing (Gann 1997). Aside from the large direct costs of financing mass screening estimated at \$12 to 18 billion per year (U.S. Preventive Services Task Force 2002, p. 128), another point of contention regarding PSA screening concerns the benefits of early detection. Most prostate cancers grow so slowly that a large majority of 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 limited success. While some studies report evidence that early detection of prostate cancer reduces disease-specific mortality, there is currently no evidence demonstrating reductions in overall mortality (Ciatto et al. 2000; Holmberg, et al. 2002; Yao and Lu-Yao 2002; Draisma et al. 2003; Concato et al. 2006). Recent randomized trials in the U.S. also find no evidence that PSA screening reduces death from prostate cancer or death from cancer in general; in fact

mortality rates were slightly higher in the group that underwent screening in Andriole et al. (2009). Against this backdrop of ambiguous benefits, the medical literature reports significant harms from prostate cancer screening which include psychological stress, needless biopsies following false positives, and overtreatment of nonlethal prostate cancers resulting in complications such as incontinence and impotence (Wang and Arnold 2002; Hawkes 2006).

Elicitation Technique for Measuring Accuracy and Consistency

We sought to construct measures of accuracy and logical inconsistency that rely on non-overlapping sets of survey items and are therefore not functionally dependent on each another. This measurement convention (basing accuracy and consistency on disjoint sets of belief elicitation) implies that any statistical associations discovered would be potentially meaningful rather than a statistical artifact of functional dependence. 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.

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?

What would you say the probability is that he will die from prostate cancer in his lifetime?

The unconditional subjective beliefs elicited by the preceding two questions are referred to as lifetime incidence and lifetime mortality, denoted $P(C \text{ Lifetime})_i$ and $P(D \text{ Lifetime})_i$, respectively. The difference between these beliefs (regarding lifetime incidence and mortality) and published point

estimates provide the basis for the individual-level measures of inaccuracy defined in the next subsection.

The interview script then continues by eliciting two conditional beliefs used to measure inconsistency of beliefs. This elicitation technique allows for infinitely many pairs of subjective conditional beliefs that are perfectly Bayesian regardless of their objective accuracy. The key feature of the elicitation script that follows is that the ratio of the conditional beliefs we elicit, but not their levels, are constrained by Bayes' Rule and the two unconditional probabilities provided. Respondents were encouraged to read the script at this point and inspect the citation to medical journals, as follows:

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*].

Given a positive PSA, what would you say is the probability that a man actually has prostate cancer?

And given cancer at the time of screening, what would you say is the probability of a positive PSA?

The resulting conditional beliefs—the probability of prostate cancer conditional on a positive PSA test, denoted $P(C|+)_i$, and the probability of a positive PSA test conditional on cancer, denoted $P(+|C)_i$ — provide the basis for measures of non-Bayesian inconsistency defined below.

As is clear from the interview script, if one applies the definition of conditional probability and then substitutes the given unconditional probabilities, the resulting restriction on the ratio of elicited conditional beliefs is as follows: $P(C|+)_i/P(+|C)_i = P(C)/P(+)$ = 2.5/5 = 1/2. Respondents may know nothing about relevant medical studies and published PSA facts. But there are infinitely many pairs of conditional beliefs that conform perfectly to the restriction imposed by Bayes' Rule regardless of whether individual i 's values of $P(C|+)_i$ and $P(+|C)_i$ are far away from published estimates of those conditional features of the PSA test. Figure 1 shows the elicited belief distributions described in this subsection with vertical segments representing published point estimates for each.

Inconsistency and Inaccuracy

We consider measures of inconsistency based on deviations of $P(C|+)_i/P(+|C)_i$ from the restriction that Bayes' Rule imposes, $P(C)/P(+)$ = 2.5/5 = 1/2. Similarly, we consider measures of inaccuracy based on the deviations $P(C \text{ Lifetime})_i / P(C \text{ Lifetime})$ and $P(D \text{ Lifetime})_i / P(D \text{ Lifetime})$, where $P(C \text{ Lifetime}) = 0.177$ and $P(D \text{ Lifetime}) = 0.028$ use the point estimates from studies cited in Table 1. Differences in levels, percentage differences and log-approximated percentage deviations using both signed and absolute deviations were analyzed.

Absolute log-approximated percentage deviations from the Bayesian restriction generate the following measure of an individual's inconsistency:⁵

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

⁵ 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 we considered 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 used in the models presented in the paper 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 later tables strengthens the case for our interpretations.

Mean absolute log-approximated percentage deviations with respect to published point estimates in Table 1 generate the following measure of an individual's inaccuracy:

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

This definition computes inaccuracy by averaging the two absolute percentage deviations of beliefs about lifetime incidence and lifetime mortality.⁶

Figure 2 presents a scatter plot of $inaccuracy_i$ and $inconsistency_i$. The bivariate regression line fit to the data in Figure 2 reveals a negative and statistically insignificant relationship between inaccuracy to inconsistency. The 24 individuals clustered along the y-axis (with $inconsistency_i = 0$) can be referred to as Perfect Bayesians in the sense that their conditional beliefs conform perfectly with the restriction imposed by Bayes' Rule. The most inaccurate individual in the sample turns out to be a Perfect Bayesian, and most others with high values of inaccuracy are well below the midpoint of the range of variation in inconsistency. Similarly, the three most inconsistent individuals in the sample have inconsistency measures below the midpoint of the range of variation for inaccuracy. These data (without conditioning on other observable features) do not suggest there is convergence of inconsistency and inaccuracy. The relationship will be analyzed further (in the next section) to investigate whether the presence of additional conditioning information reveals any positive relationships associating inconsistency with inaccuracy.

Accuracy Contrasts among Perfect Bayesians and Other Subsamples Based on Discretized Inaccuracy

Table 3 presents four binary contrasts of mean inaccuracy (interpreted as the log-approximated percentage deviation from published point estimates) broken out by subsamples of those with

⁶ 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 deviations alone, and an average of five deviations based on all five elicited beliefs, revealing no substantive changes.

consistent versus inconsistent beliefs. The four mean contrasts and corresponding t statistics are, of course, not independent because they use the same observations of inconsistency discretized by different thresholds. Consistent versus inconsistent groupings in Table 3 are as follows: Perfect Bayesians (*inconsistency* = 0) versus non-Bayesians (*inconsistency* > 0); below-median versus above-median *inconsistency*; bottom versus upper quartiles of *inconsistency*; and Near Bayesians (an inclusive classification for anyone whose inconsistencies can be modeled as Bayesian beliefs plus a noise term) versus so-called Emersonians (explained below) comprised of those who commit gross errors in conditional probabilistic reasoning.⁷

The rows of Table 3 contain mean values of: inaccuracy, signed inaccuracy, four log deviations of elicited beliefs, inconsistency and signed inconsistency. Signed inconsistency removes the absolute values from the definition given previously, which results in deviations with different signs cancelling (to some extent) when summing over individuals. Reading horizontally across the first row, Table 3 indicates that the average Perfect Bayesian (recall that 24 individuals have *inconsistency*_{*i*} = 0) are more inaccurate than the rest of the sample (1.26 versus 0.90). Similarly, the lower half of the inconsistency distribution has greater inaccuracy than the upper half (1.08 versus 0.87), and the lower quartile has greater inaccuracy than the upper quartile (1.26 versus 0.77). According to the fourth cut of the sample into Near Bayesians and Emersonians, accuracy is, once again, negatively associated with consistency: mean inaccuracy of 1.08 among the consistent versus 0.78 among the inconsistent.

The second row of Table 3 shows that the beliefs of consistent respondents tend to be too small, whereas beliefs of inconsistent individuals tend to overshoot with respect to the point estimates published in medical journals. Consistent individuals' beliefs are not, however, generally any closer to the published frequencies. Rows 3 and 4 show mean log deviations for lifetime incidence and

⁷ The label is inspired by Emerson's "Self Reliance" (1841) 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."

mortality, which breaks out the two components averaged in *signed* inaccuracy and inaccuracy. These disaggregated bivariate contrasts in Table 3 show no general tendency for consistent individuals to have more accurate beliefs regardless of which threshold is used to dichotomize the sample, confirming the absence of any positive association in the scatterplot in Figure 2.

Taxonomy of Inconsistencies: Emersonians and Near 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 below) and could therefore be modeled as if they were generating Bayesian beliefs with an error term (i.e., only moderately inconsistent conditional beliefs). Other subjects' beliefs involve more basic violations of probabilistic logic. The former group is referred to as Near Bayesians and the gross violators are referred to as Emersonians.

We define three types of gross violations of probabilistic reasoning, any one of which would indicate a belief generating process 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 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 $\frac{1}{2}$:

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

Elicited probabilities 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 in for both left- and right-hand-side terms, the numerators of those

conditional probabilities are of course the same while the denominators have unknown values. But $P(C)=0.025 < P(+PSA)=0.05$, which implies $P(C|+) \leq P(+|C)$. This inequality 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 from probabilistic reasoning.

The third logical error is $P(C|+)_i = P(+|C)_i$. Given the information provided to respondents that explicitly mentioned 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 Is Associated with Economic Losses?

Two Channels For Inconsistency to Generate Losses

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 , interpreted as accounting for all inter-personal differences aside from inconsistency and beliefs. Let ω represent states of nature drawn from a standard probability measure on the universe Ω . In the PSA testing context, states of nature can be thought of as terminal nodes on a large probabilistic event tree leading to contingencies consisting of combinations of the following events: cancer, positive PSA, types of prostate cancer, and treatment options in the event of cancer.

States of nature include contingencies with successful treatments, partially successful treatments (i.e., with costly side-effects), and contingencies with unnecessary surgeries (e.g., surgery removing a slow-growing cancer that would not have proved lethal if left untreated). States of nature

include other contingencies in which valuable treatment options are missed. From the root node of the event tree, the first step has two branches corresponding to the unobserved events of prostate cancer and no prostate cancer. The second step has two branches, +PSA and -PSA, stemming from the cancer node and the same two from the no cancer node.

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 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 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 → surgery; or biopsy that successfully rules out prostate cancer.

From this long yet far-from-complete list of states ω , numerous value judgments specific to each individual go into assigning payoffs which, in turn, provide a ranking of all terminal nodes. For example, as the 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 while others will prefer to live more years but with side effects from treatment. Still others might prefer to never be tested or diagnosed regardless of their 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. In other words, even when $b_i = b_{i'}$ and $\delta_i = \delta_{i'}$, individuals i and i' may nevertheless assign different losses to each contingency ω and possibly make different *ex ante* loss-minimizing decisions about PSA testing, $t_i^* \neq t_{i'}^*$, whenever $\theta_i \neq \theta_{i'}$. This allows for full heterogeneity across people in ranking the PSA-related contingent outcomes and does not presume there is a

universally correct decision (i.e., to test, or not to test).

Given these definitions, we now assume there is a standard stochastic structure generating ω conditional on subjective beliefs b_i summarized by the conditional pdf $f_{\omega|b}(\omega, b_i(\delta_i))$. To organize our thinking about how inconsistency could, in theory, generate economic losses and derive testable predictions, we assume that the loss function L depends on states of nature, inconsistency (exerting a direct effect net of its effect on subjective beliefs), the testing decision denoted t_i , and person-specific parameters needed to rank all contingencies: $L = L(\omega, \delta_i, t_i; \theta_i)$. Taking beliefs and inconsistency as given, the decision maker computes expected loss (i.e., risk) at each element in the binary choice set ($t_i = 0$, which codes the decision not to have a PSA test, or $t_i = 1$, which codes the decision to have a PSA). The decision maker computes the following two expected losses and chooses the minimum of the two:

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

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

The optimal PSA testing decision is to choose t_i to minimize risk: $t_i^* = \operatorname{argmin}_{t_i \in \{0, 1\}} R_{t_i}$.

This model of expected loss minimization implies there are two theoretical channels through which inconsistency (δ_i) can influence PSA decisions (t_i^*). Inconsistency must influence accuracy of beliefs directly or influence the mapping from beliefs into PSA decisions. The empirical strategy is therefore to examine these two channels separately after linearizing the functional dependence of b_i on δ_i and of t_i^* on δ_i . Thus, the regressions in the following two subsections seek to detect effects of δ_i on b_i and δ_i on t_i^* , respectively. A necessary condition for *inconsistency* to lead to losses is for it to have a statistical association on at least one of these two channels.

Empirical Model Allowing for Conditional Effects of Inconsistency on Inaccuracy

Table 4 shows results from a regression of inaccuracy on inconsistency with a full set of controls: having consulted written information, information processing (i.e., weighing pros and cons),

social influencers, a quadratic function of age, other individual characteristics from the survey, and subfield indicators. Comparing the bivariate regression line in Figure 2 (where the regression coefficient on inconsistency was $-.06$) to the conditional model in Table 4 (where the coefficient on inconsistency is nearly the same, changing only to $-.08$), one finds 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 see a statistically significant and positive coefficient suggesting a positive association between consistency and accuracy of beliefs.

There are several puzzling effects in Table 4. Consulting written information paradoxically increases inaccuracy of beliefs. On the other hand, the variable “weighing pros and cons” appears to have a beneficial effect reducing inaccuracy but 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.

Empirical Models Allowing for Conditional Effects of Inconsistency on PSA Testing

Table 5 presents estimates of four linear probability models with t statistics computed using robust standard errors.⁸ The *fundamental model* assumes PSA decisions are a function of the five subjective beliefs (proxying for beliefs about risks of cancer and net benefits of the PSA test) and a quadratic function of age. The *add-info-processing model* assumes that PSA decisions are a function of

⁸ Logit and probit models produce qualitatively identical results and are available from the authors upon request. Similar 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.

everything appearing in the *fundamental model* and information acquisition, information processing, and inconsistency. Finally, the *add influencers model* (again encompassing previous models) allows the probability of taking a PSA to depend on social influencers. The first three models use self-reported PSA testing decision as the dependent variable and the fourth model uses PSA recommendation as the dependent variable to investigate whether adding conditional information can explain the large unconditional gap between PSA taking and recommendation, 46 versus 91 percent, respectively.

We find statistical confirmation of self-reports that most economists do not weigh costs and benefits when deciding whether to have a PSA test. The joint test that the first five regressors measuring risk belief have zero coefficients 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 corresponding to tests of that hypothesis, revealing weak joint explanatory power of subjective beliefs in the first two models. This weak explanatory 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, subjective costs and benefits about PSA testing begin to have statistically significant explanatory power once information about social influences is included. Across all three models, however, individual beliefs have surprisingly weak effects on the probability of having PSA testing. For example, the perceived risk of incontinence, which one might have guessed would strongly condition men's evaluations of the test's desirability, has very modest effects, implying that a man whose perceives risk of incontinence to be twice as large as the average perceived risk is, at most, 6 to 8 percentage points less likely to have a PSA. Coefficients on information acquisition and processing (i.e., pros-and-cons deliberation and logical inconsistency) are nowhere large or statistically significant.

The *doctor influenced* variable reveals strong conditional correlation between reliance on a

doctor's recommendation and taking the PSA test despite the obvious incentive mismatch in doctor-patient interaction leading to well-documented problems of defensive medicine, over-diagnosis, over-prescription, over-treatment and other potential problems that economists are well aware of (see Studdert et al. 2005; Behrens, Güth, Kliemt and Levati 2005; Loewenstein 2005; and Sorum et al. 2004, for more on doctor-patient incentive mismatch).

PSA Recommendation

The pairwise correlation between PSA recommendations and self-reported decisions is a surprisingly small 0.09 and far from statistical significance. As mentioned above, the unconditional rate of recommendation is double the rate of PSA test taking. The fourth model in Table 5 therefore investigates whether inclusion of conditioning information shrinks this gap or reveals any association between logically inconsistent beliefs and the PSA recommendations. To keep the sample size the same as the three PSA decision models in Table 5, 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: 46 versus 85 percent. Beliefs about costs and benefits have more predictive power for PSA recommendations than for PSA decisions. But once again, inconsistency of beliefs plays a very limited role.

Section 4: Discussion

Summary

The primary goal of this paper was to elicit subjective belief data that provide separate measures of consistency with respect to Bayes' Rule and accuracy with respect to objective frequencies. These measures of inconsistency and accuracy revealed new empirical evidence regarding the divergence, or lack of positive association, between two importantly distinct normative criteria for

evaluating belief data that social scientists frequently assume are jointly present when characterizing rational decision making. This presents a challenge to those who would attempt to characterize subjective beliefs along a univariate spectrum of rationality.

Our elicitation technique for belief inconsistency announces two unconditional probabilities and then elicits conditional probabilities whose ratio is constrained by the definition of conditional probability. Individuals (even economists who are well equipped to compute Bayes' Rule) vary considerably in the extent to which their conditional beliefs conform to restrictions imposed by the logic of probability theory.

A second goal of this paper was to organize the elicited belief data in a manner that tests for evidence that inconsistent beliefs cause economic losses, which must take place through two channels linking inconsistency to losses in the expected loss minimization framework introduced at the beginning of Section 3. The data show no evidence that inconsistent beliefs are associated with impaired accuracy or action rules mapping beliefs into decisions about PSA testing. Subjective beliefs about risks of prostate cancer and PSA testing are jointly insignificant until the role of social influences is included in regression models, whether linear probability or logit models are estimated. The absence of any positive association between consistency and accuracy shows up in numerous alternative measures of inconsistency and inaccuracy based on percentage-point deviations, percent deviations, and log-approximated percent deviations regardless of whether signed or absolute deviations are used.

With the usual caveats required when 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 when deciding whether to have a PSA test. More than half said they had not weighed pros and cons. One may ask whether these data are simply too noisy to reveal underlying mechanisms that might reveal positive associations between consistency and accuracy. 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 the simple heuristic of following doctors' advice, sometimes referred to as a *white-coat heuristic*: If you see a white coat when in the hospital or at a doctor's office, then do what it says. The social influencer indicator variables, especially *doctor influenced*, add significant explanatory power to the conditional models. Whether trusting one's doctor is effective in any broader normative sense is not addressed by our findings.

We can, however, most likely rule out one interpretation, namely, that relying solely on the doctor's advice can be rationalized within the framework of expected payoff maximization. There is abundant evidence of incentive mismatch between doctors and patients that leads to defensive medicine (i.e., treatments provided for the doctor's benefit of legal protection) and overtreatment of cancers that would never have caused any patient's death. There is also abundant evidence documenting large gaps in doctors' statistical literacy and knowledge of research and relevant statistical evidence. On the other hand, the time costs to access information about prevalence and mortality of prostate cancer, together with evidence-based recommendations on screening and treatment, would amount to little more than a few mouse clicks as this information is readily available online (e.g., the U.S. Preventive Services Task Force online database). Moreover, an expected utility maximization model that leads to the action rule of relying on the doctor's advice without conditioning on other sources of information would have to take a very restrictive form for patients' subjective beliefs about risks of cancer and PSA testing not to influence the propensity to take the PSA test.

Interpretations

Why would 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. According to

them (and others), non-standard reasoning processes that generate behavior inconsistent with axioms of internal consistency can be defended and, in some contexts, shown to have advantages over decision processes adhering strictly to consistency (e.g., Gilboa and Schmeidler, 1995; Samuelson, 2001; Aragonés et al., 2005; Spiegel, Heifetz and Shannon, 2007; Robson and Samuelson, 2009; Bardsley et al. 2010). Grunwald and Halpern (2004) identify a related problem in which non-Bayesian updating provides more precise predictions. Less-is-more effects, where non-standard behavioral beliefs provide benefits in terms of improved predictive accuracy by ignoring information also appears in both theoretical and empirical studies (e.g., Hogarth and Karelia, 2005, 2006; Baucells, Carrasco and Hogarth, 2008; Berg and Hoffrage, 2008; Goldstein and Gigerenzer, 2009).

Sugden (1991) argues against the normative interpretation of expected utility theory. Starmer (2000, 2005, 2009) provides illuminating historical and methodological analysis of normative debates about Bayesian reasoning and expected utility theory arriving at similar conclusions. 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 decision makers' natural health decision-making environments) which raises questions about whether economists should assume that inconsistency will be competed out of existence or reduced as the result of accumulating more experience (c.f., Braga et al 2009). In high-stakes decisions (e.g., medical decisions with substantial mortality risk, financial decisions involving a large fraction of one's wealth, or career and relationship advice among loved ones), many of us routinely apply normative criteria that clearly go beyond internal consistency.⁹

⁹ 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 decision scientist who was considering the offer deliberately chose to deviate from the normative theory which he knew well. When colleagues asked him why he didn't just choose a prior, compute expected utilities associated with each of his options and choose the action with maximal expected payoff, the decision theorist expressed exasperation at the insufficiency of EU maximization for making such a high stakes decision: "Come on, this is serious!" he said (Gigerenzer, 2004, p. 62). This anecdote illustrates that even those who best understand the mechanics of Bayesian reasoning routinely find it inapplicable as a decision-making tool.

Decision Making Process in PSA Testing

Table 1 showed that only 46 out of 128 respondents reported having weighed pros and cons when deciding on PSA testing. Among those who reported not weighing pros and cons were 16 who did not do so despite having reported perceiving potential harms. This suggests a thought process different from those typically assumed in economics, which would seem to find support in Gilboa, Postlewaite and Schmeidler's (2009, p. 285) “view of rationality that requires a compromise between internal coherence and justification....” An important justification in the social and family environments in which people commonly make important medical decisions is provided by social influencers (e.g., having the PSA test because my wife or doctor told me to do so, or because someone I know did so). One of the most frequently encountered non-standard decision procedures in evolutionary models is imitation and, more generally, conditioning action on social cues (Boyd and Richerson, 1985).

Guess-50 Heuristic

Respondents may have simply guessed 50 percent when facing belief elicitation about which they had no strong prior. We coded the number of times respondents guessed “50 percent” to see if completely uninformed priors leading to a *guess-50 heuristic* was correlated with either *consistency* or *accuracy*. Among the five elicited beliefs, the maximum number of times anyone in the sample guessed 50 is twice. Interestingly, those 22 individuals who guessed 50 two times had more accurate beliefs (mean inaccuracy of 0.71, $sd = 0.01$) than those who never guessed 50 (mean inaccuracy 1.02, $sd = 0.09$). Two of the 24 Perfect Bayesians guessed 50 two times. Emersonians and Near Bayesians guessed 50 at roughly the same rates. And inconsistency was uncorrelated with guessing 50.

Additional Evidence Regarding Social Influences on PSA Decisions

There 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. Respondents who weighed pros and cons were PSA test takers at a rate of 76 percent and those who did not weigh pros and cons were PSA takers at a rate of 61 percent, only a modest 15 percentage point difference. This difference disappears entirely in the subsample of those who reported being influenced by at least one other person (most commonly, a spouse). This suggests that social influence can completely override cost-benefit thinking. This difference complements evidence from the regressions pointing to the importance of social influencers.

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 was 120. 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).

Conclusion

The conclusions we draw are not categorically in conflict with the possibility of real-world benefits from adhering to axioms of logical consistency. Rather our empirical study of distinct normative criteria for evaluating subjective beliefs demonstrate the potential for surprising divergence and need for larger consideration of multiple normative criteria (c.f., Berg, 2003). It would seem that matching particular normative criteria to the reward structure inherent in specific decision-making contexts might improve normative analysis over universal claims that rational decision makers must conform to certain axioms across all decision domains. Economists are presumably as familiar with the normative benchmarks of consistency and accuracy as anyone, yet they vary substantially in the degree to which their subjective beliefs adhere to the consistency requirements of probabilistic logic, the accuracy of their beliefs, and in the medical decisions they make. Despite this substantial variation, no positive associations between inconsistency and inaccuracy were observed and, more than occasionally, the association seemed to go in the opposite direction. This result conflicts with standard methodological norms that tacitly assume convergence among multiple normative criteria. The data support the view articulated in Gilboa, Postlewaite and Schmeidler (2009, p. 288), who 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.

References

- 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.
- 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, Nathan, and Ulrich Hoffrage, "Rational Ignoring with Unbounded Cognitive Capacity," *Journal of Economic Psychology*, 29(6) (2008), 792-809.
- 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.
- 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.
- 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.
- 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.
- Draisma, Gerrit, Rob Boer, Suzie J. Otto, Ingrid W. van der Crujisen, 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).
- 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 in 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-Axelsson, 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.
- 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, "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.

- Lusardi, A. and O.S. Mitchell, "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy And Retirement Readiness," Working Paper (2009), Dartmouth University.
- 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.
- Robson, A., and L. Samuelson, L., "The Evolution Of Time Preference With Aggregate Uncertainty," *American Economic Review*, 99 (2009), 1925-1953.
- Samuelson, L., "Analogies, Adaptation, And Anomalies," *Journal of Economic Theory*, 97 (2001), 320-367.
- Savage, L., *The Foundations of Statistics*, (New York, NY: Wiley, 1954).
- 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).
- 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. Heifetz, 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.
- Tversky, Amos and Kahneman, Daniel, "Judgment Under Uncertainty: Heuristics And Biases," *Science*, 185 (1974), 1124-1131.
- 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.
- 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

	<u>Fraction Yes</u>	<u>Number of Valid Responses</u>		
<i>individual characteristics</i>				
Keep \$3 cash?	0.12	133		
Give \$3 to charity?	0.71	133		
Inaccuracy measures based on absolute log Economist?	0.17	133		
Work is applied as opposed to theoretical?	0.92	133		
Neoclassical methodological orientation?*	0.75	124		
50 years old or older**	0.75	128		
	0.62	133		
<i>PSA decision and recommendation</i>				
Did you have a PSA?	0.46	133		
Would you recommend a PSA to men in their 50s?	0.91	124		
<i>information acquisition, perceived harms, and mode of information processing</i>				
Written info?	0.22	131		
Medical journal?	0.05	131		
Harms?	0.25	122		
Weighed pros and cons?	0.36	128		
<i>social influences</i>				
Doctor influenced?	0.58	133		
Spouse or relative influenced?	0.07	133		
Nobody influenced?	0.15	133		
	<u>Mean</u>	<u>Std</u>	<u>Published</u>	
	<u>Elicited</u>	<u>Dev of</u>	<u>Number of</u>	<u>point-</u>
<i>elicited frequencies</i>	<u>Value</u>	<u>Mean</u>	<u>Responses</u>	<u>estimates***</u>
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

*Primary subfield specializations were collected, too: 7 percent econometrics, 12 percent finance, 5 percent health economics, 7 percent economic history, 5 percent 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. Mean self-reported age was 51 years old, with a strong majority (119) reporting ages of 40 or older. ***Stanford et al's (1999) NCI SEER study and Harris and Lohr (2002).

Table 2: Medical literature arguing against automatic PSA screening of asymptomatic men

<u>Journal</u>	<u>Author(s)</u>	<u>Comment</u>
<i>Archive of Internal Medicine</i>	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."
<i>Annals of Internal Medicine</i>	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..."
<i>Journal of the National Cancer Institute</i>	Draisma et al (2003)	"Whether asymptomatic men benefit from screening for prostate cancer is an unresolved question."
<i>New England Journal of Medicine</i>	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."
<i>European Journal of Cancer</i>	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 (2003) in the <i>British Medical Journal</i>]
<i>American College of Physicians</i>	Concato (1999)	"Routine PSA measurement without a frank discussion of the issues involved is inappropriate."
<i>Epidemiology</i>	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 (1995) in <i>JAMA</i>]
<i>Journal of the American Medical Association (JAMA)</i>	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."

*A common recommendation of studies raising questions about PSA screening of asymptomatic men is that doctors should provide patients with information regarding pros and cons while encouraging patients to decide about PSA testing on an individual basis. Medical communication experts refer to this as the balance-sheet approach, with the goal of asking 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 (part 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 have a policy of automatic screening based on age following recommendations supportive of automatic PSA screening by the American Cancer Society and the American Urological Association.

Table 3: Contrasts between consistent and inconsistent subsamples

	consistent	inconsistent		consistent	inconsistent		consistent	inconsistent		consistent	inconsistent		consistent	inconsistent
				<u>60</u>			<u>36 weakly</u>	<u>34 weakly</u>						
				<u>strictly</u>	<u>65 weakly</u>		<u>below</u>	<u>above</u>						
		<u>101</u>		<u>below</u>	<u>above</u>		<u>25th</u>	<u>75th</u>						
		<u>Deviators</u>		<u>median</u>	<u>median</u>		<u>percentile</u>	<u>percentile</u>						
<u>Pooled</u>	<u>24 Perfect</u>	<u>from</u>	<u>t</u>	<u>incon-</u>	<u>incon-</u>	<u>t</u>	<u>incon-</u>	<u>incon-</u>	<u>t</u>	<u>80 Near</u>	<u>45</u>	<u>t</u>	<u>Bayesians</u>	<u>Emersonians</u>
<u>Mean</u>	<u>Bayesians</u>	<u>Bayes</u>	<u>stat</u>	<u>sistency</u>	<u>sistency</u>	<u>stat</u>	<u>sistency</u>	<u>sistency</u>	<u>stat</u>	<u>Bayesians</u>	<u>Emersonians</u>	<u>stat</u>	<u>Bayesians</u>	<u>Emersonians</u>
<i>Inaccuracy measures based on absolute log deviations (and signed log deviations) of elicited beliefs</i>														
inaccuracy	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	
<i>Log deviations</i>														
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	
<i>Measures of inconsistency</i>														
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 percentage error of the elicited ratio, *sensitivity/posterior*, relative to the correct ratio of 2. Signed inconsistency is the same as inconsistency but without absolute values.

Table 4: Regression of inaccuracy

<u>predictors</u>	<u>coef</u>	<u>t</u>
inconsistency	-0.08	-0.6
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
constant	3.11	2.0
R2	<i>0.24</i>	
Sample Size	<i>117</i>	

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

Three PSA-Decision Prediction Models:

predictors	<i>fundamental</i>		<i>add info- processing</i>		<i>add influencers</i>		PSA- Recommendation (1/0)	
	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	<i>0.34</i>		<i>0.38</i>		<i>0.46</i>		<i>0.18</i>	
Pr(test stat>observed H0)	<i>0.13</i>		<i>0.14</i>		<i>0.03</i>		<i>0.01</i>	
Sample Size	<i>121</i>		<i>114</i>		<i>114</i>		<i>114</i>	

*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.

Figure 1: Elicited subjective belief distributions

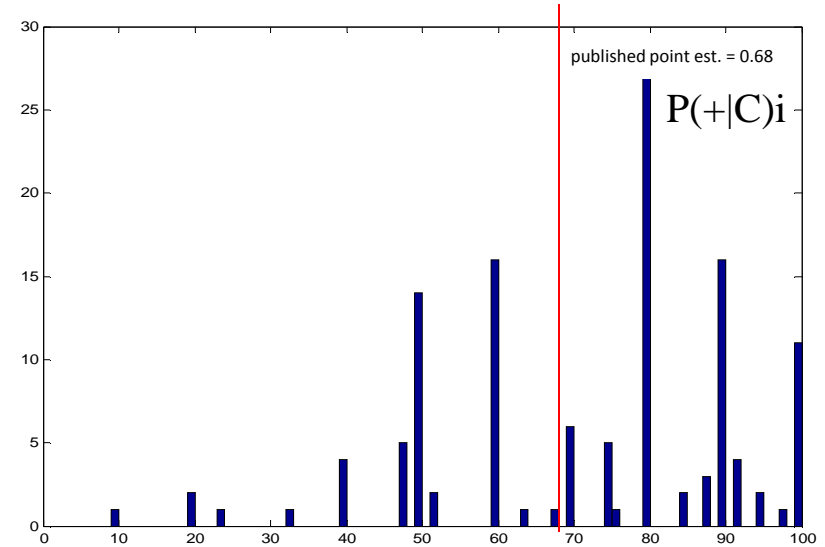
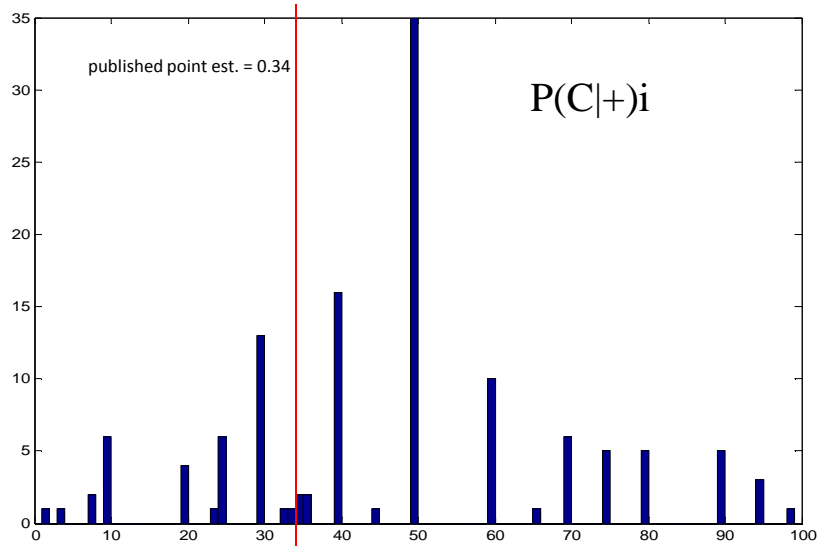
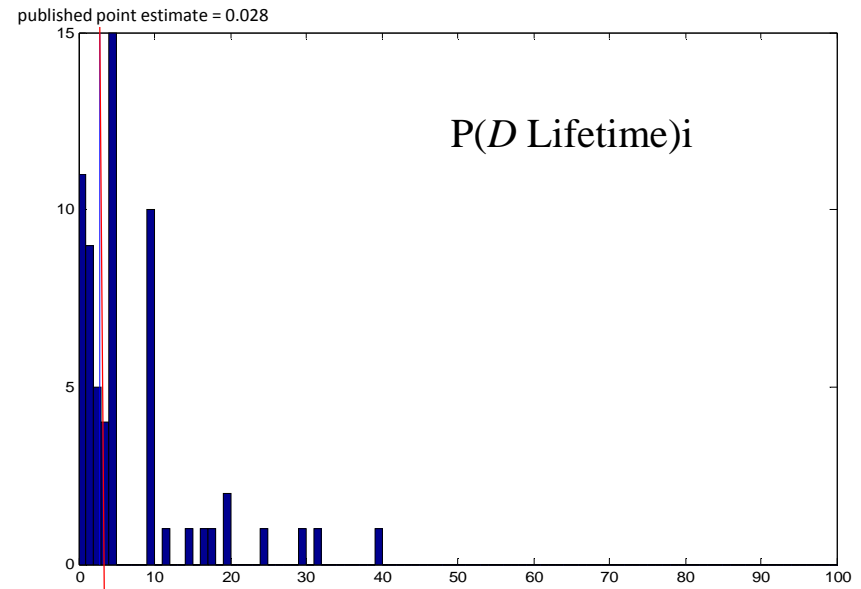
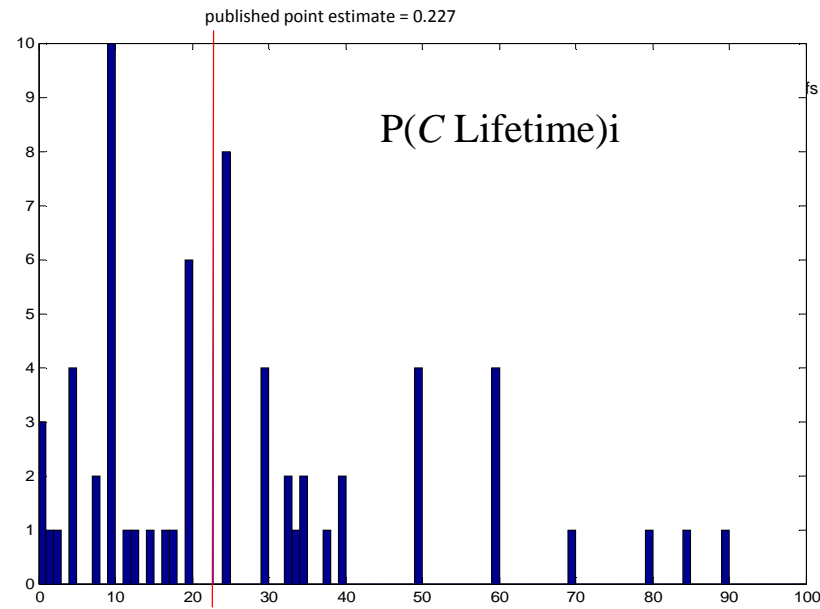
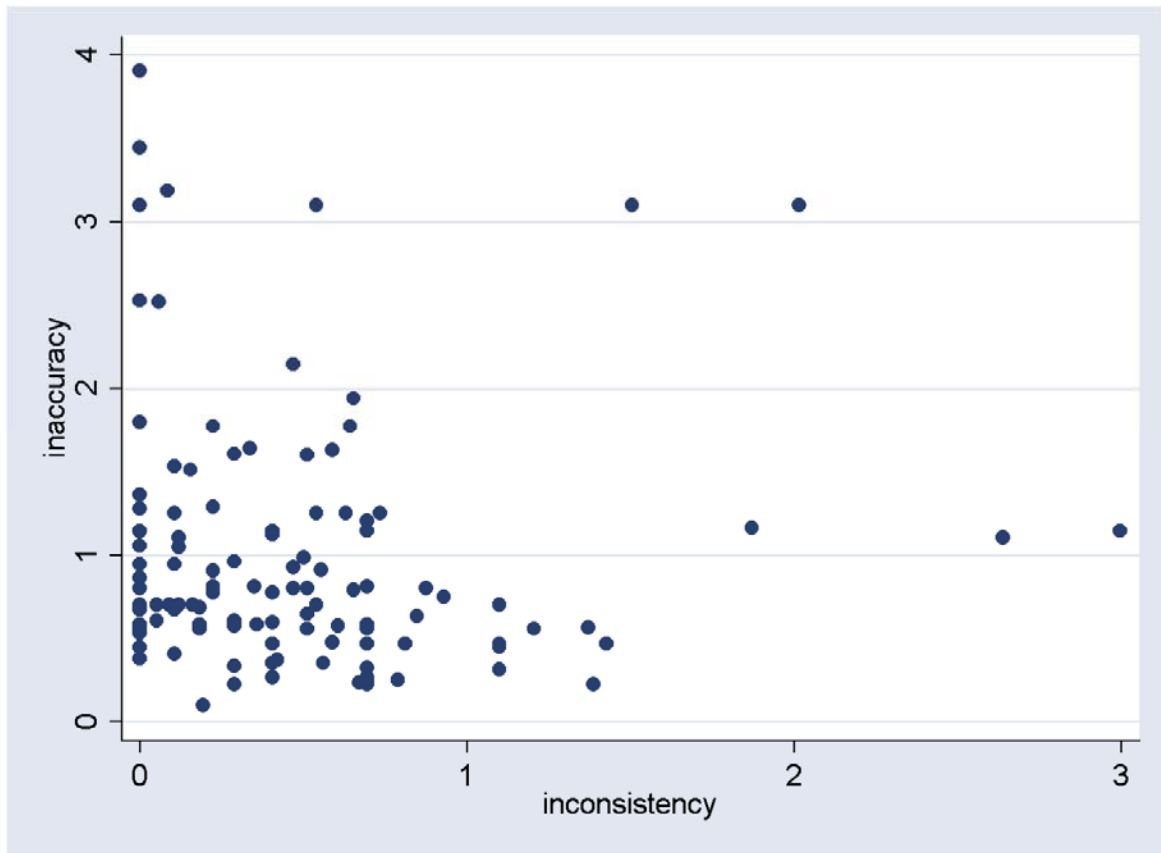


Figure 2: Inconsistency versus inaccuracy (N=125)



*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.