Consistency versus accuracy of beliefs: Economists surveyed about PSA

Nathan Berg
prof.berg@gmail.com

Guido Biele
biele@mpib-berlin.mpg.de

Gerd Gigerenzer
gigerenzer@mpib-berlin.mpg.de

Abstract: *We surveyed 133 attendees at the 2006 meeting of the American Economic Association (AEA) about their beliefs and behavior regarding the Prostate Specific Antigen (PSA) test for prostate cancer. Accuracy of beliefs was measured by comparing elicited rates of lifetime incidence and mortality of prostate cancer with point estimates published in medical journals. Logical consistency was measured in terms of percentage deviations from a restriction imposed by Bayes Rule on beliefs about two related conditional probabilities—the sensitivity of the PSA test and the posterior probability of cancer given a positive PSA. Contrary to the hypothesis that individuals can be ordered along a uni-dimensional spectrum of rationality, which would require positive within-person correlation among the extents to which an individual adheres to distinct normative criteria, our data reveal negative or no correlation between consistency and accuracy of beliefs. Social influences are much stronger predictors of PSA decisions than beliefs about PSA testing, beliefs about cancer risks, and logical consistency.*

Key words: logical consistency, accuracy, non-Bayesian beliefs

Consistency versus accuracy of beliefs: Economists surveyed about PSA

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. –Ralph Waldo Emerson, from "Self Reliance," 1841

Section 1: Introduction

This paper focuses on an empirical measure of consistency of beliefs with respect to Bayes Rule. We investigate how consistency correlates with accuracy of beliefs and whether consistency predicts behavior. The empirical results draw on survey data collected from attendees at the 2006 Allied Social Science Associations meetings in Boston, where we elicited men's beliefs about the risks of prostate cancer and the statistical properties of the Prostate Specific Antigen (PSA) test, a blood test widely used to screen for prostate cancer.

The medical literature documents that PSA testing, when used as a screening device for asymptomatic men, may have economically significant costs in the form of medical complications, psychological stress from false positives, and needless treatment of nonlethal cancers (Stanford et al, 1999; U.S. Preventive Services Task Force, 2002). The National Cancer Institute (www.cancer.gov) and the U.S. Preventive Services Task Force (www.ahrq.gov) explicitly recommend that, instead of undergoing automatic PSA screening, men should weigh costs and benefits on an individual basis before having the test. According to these sources, the decision to have a PSA test is a high-stakes choice that merits deliberation about trade offs. Thus, according to the standard cost-benefit model of individual decision making, the PSA decision should be sensitive to beliefs about risks and statistical properties of the PSA test. The purpose of this study is to investigate whether inconsistent beliefs are associated with greater-than-average distortions in the accuracy of beliefs or systematically different behavior—a necessary condition for logical inconsistency to cause economic harm.

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). Based on the extensive literature on non-Bayesian behavior, one might presume that deviations from Bayesian norms have important economic consequences. However, there is little evidence from the field[1] showing that non-Bayesian beliefs incur significant economic costs.

Looking for evidence regarding costs of inconsistency, we test a model that assumes that PSA decisions are based on minimization of a loss function which depends on beliefs about prostate cancer risks, beliefs about the quality of the screening instrument, and the PSA decision itself. We do not attempt to estimate the parameters in this loss function, which are likely to vary across individuals (e.g., weights assigned to pairs of actions and states of nature, which would include whether the subject has lethal cancer and the possible presence of side effects from treatments chosen). We argue that economic losses from inconsistent beliefs in the context of deciding whether to have a PSA test must show up through one of two possible channels: either in systematically different assessments of risks and benefits, or in different PSA decision outcomes holding beliefs about risks and benefits constant.

One form of evidence consistent with the claim that logical inconsistency has economic effects would be the finding that logical inconsistency correlates with the objective inaccuracy of beliefs about risks and benefits. A second form of evidence for economic losses resulting from inconsistency would be that inconsistency predicts actual PSA decisions, net of its influence on beliefs about the frequencies used to assess costs and benefits. Assuming an informational loss function that depends on deviations between subjective and objective frequency estimates, the first form of correlation is necessary for showing informational losses associated with inconsistency. The second form of correlation is necessary for showing an outcomes-based channel for inconsistency to affect the loss function. The intuition is simple: if inconsistent beliefs are costly, then one expects that inconsistent men have either less accurate beliefs or systematically different behavior conditional on beliefs (i.e., different mappings from any given combination of perceived costs and benefits into PSA decisions). Because accuracy and consistency measures are computed from separate sample items referring to different probabilities—lifetime risks of prostate cancer on the one hand, and point-in-time properties of the PSA test on the other—the elicitation scheme allows for variation in the consistency of beliefs about PSA while holding accuracy of perceived risks constant.

Raising questions about whether deviations from standard normative benchmarks are individually and socially costly (or perhaps even beneficial) does not imply broader skepticism about the abundant

[1] Distinctions as to what constitutes experiments in the field have been put forward by Harrison and List, (2004), who argue that the term “field experiment” should be reserved for cases in which subjects are observed in their natural settings.

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 price?

More generally, one may inquire into the evidential basis justifying the role of logical consistency in economic models. Virtually every model with utility functions, for example, depends on transitivity axioms, often defended as normatively appealing based on inconsistent individuals' theoretical vulnerability to money-pump exploitation (Davidson, McKinsey and Suppes 1955, p. 146; Raiffa 1968, p. 78).[2] Although the existence of transitivity violations is by now beyond doubt (Tversky, 1969; Grether and Plott, 1979; 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.[3] The Savage axioms underlying expected utility theory are another example of consistency criteria whose normative status is widely accepted despite a lack of evidence demonstrating that deviators suffer significant losses.[4] Recent empirical and theoretical work demonstrating large social components in the formation of beliefs (Di Tella, Galiani and Schargrodsky, 2007) and their role in generating aggregation reversals (Glaeser and Sacerdote, 2007) provide motivation for our empirical approach to describing decision processes as a function of social influences, without assuming that deviations from standard decision-making axioms necessarily lead to social inefficiency.

To improve the chances of finding empirical links between logical consistency, the objective accuracy of beliefs, and PSA decision outcomes, the data reported in this paper were collected primarily from economists. Previous studies have shown that economists behave differently from noneconomists

[2] Exceptions include the models of Rubinstein and Spiegler (2006), Laibson and Yariv, (2004), 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, 1989) 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.

[3] 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. 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, raising questions about whether economists should assume that inconsistency will be exploited in competition and mitigated by experience.

[4] 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, 2004) for more on expected utility as a normative theory.

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 noneconomists, 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 noneconomists 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 those of the general population, whose poor understanding of statistics and "health literacy" has been well documented (Williams et al, 1995; Baker et al, 1998; Parker et al, 1995).

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. In fact, 23 of 123 economists in the sample (almost 19 percent) and 1 out of 10 noneconomists conformed perfectly to Bayes Rule. The remaining 109 in our sample of 133 violated Bayes Rule to varying degrees. 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, which, in turn, helps reveal what role—if any—logical consistency plays.

Section 2 describes how the data were collected and reports descriptive statistics. Section 3 presents the main findings in the form of regressions that measure the predictive power of logical consistency on two dependent variables, the objective accuracy of beliefs and the PSA decision. Section 4 investigates the robustness of these findings by exploring alternative measures of inconsistency based on alternative survey items. Section 5 discusses possible interpretations of these results and considers further evidence that social influences are an order of magnitude more important than subjective costs and benefits in explaining whether respondents choose to have a PSA test.

Section 2: Description of data

We surveyed 133 attendees of the annual meeting of the American Economic Association (AEA), also known as the Allied Social Science Associations meetings, which lasted for three days, January 6-8, 2006, in Boston, Massachusetts. The meeting is regularly attended by approximately 9,000 economists and is the primary location for initial rounds of the job market for economists. Although noneconomists attend the meetings, attendees are predominantly economists. Our sample contained 123 economists and 10 noneconomists, most of whom were political scientists or academics working in related fields.

Surveys were conducted as face-to-face interviews based on a written script. The script was visible to respondents and the interviewer encouraged respondents to read any sample items for themselves when asked for clarifications. The interviewer adhered closely to the script, which is reproduced in Appendix 1, with interviews lasting three to 10 minutes. No time limit was imposed. Most interviews were collected a few meters from the registration desk at the AEA meetings, which 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 potential survey respondents with a memorized introductory statement offering respondents a choice of $3 or a Swiss chocolate bar, and assurances that the survey would be short. Survey respondents who chose $3 over chocolate were asked if they wanted to donate this participation fee to charity, which a majority did. Table 1 contains basic survey results.

Almost half of respondents, 46 percent, reported having had a PSA—65 percent among those 50 and older. Most respondents (85 percent) said they would recommend that an asymptomatic man in his 50s should have a PSA, with almost no difference in recommendation rates by age. The column under the heading “Valid Observations” in Table 1 shows that item non-response was an issue for several questions, although not the ones we would have expected. Five respondents refused to provide responses for *posterior probability* (defined below). Seven refused to provide responses for *sensitivity*. Nine refused to recommend whether a man in his 50s should have a PSA. And nine refused to classify their work as either “more applied” or “more theoretical.” None refused to say whether he had taken a PSA.

Elicited Frequencies

The survey elicited the following five frequencies concerning prostate cancer risks and the quality of the PSA test:

- *lifetime incidence* (the probability that a randomly drawn male in the U.S. is diagnosed with prostate cancer within his lifetime) denoted Pr(C)

- *lifetime mortality* (the probability that a randomly drawn male in the U.S. dies of prostate cancer within his lifetime) denoted Pr(D)
- *incontinence probability* (the probability of incontinence conditional on surgical treatment for prostate cancer) denoted Pr(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 Pr(C | + PSA)
- *sensitivity*[5] (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 Pr(+PSA | C)

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, these five belief variables should have joint explanatory power as a proxy for perceived net benefits of testing.

Consistency and Accuracy of Beliefs

We sought to construct a measure of logical inconsistency that does not depend on the accuracy of stated beliefs. To accomplish this, our elicitation scheme allowed for many pairs of belief responses that are perfectly consistent with Bayes Rule regardless of their accuracy. We provided survey respondents with two point estimates published in the *Annals of Internal Medicine* (Harris and Lohr, 2002): the probability that an asymptomatic U.S. male in his 50s screened for the first time will have a positive PSA test, P(+PSA) = 0.05; and the single-point-in-time probability of prostate cancer in the same population, P(C)= .025. These are unconditional probabilities. Accordingly, the interview script emphasized the fact that P(+PSA) includes both true and false positives, and that P(C) includes both cancers that are detected by the PSA test and those that are not.

Having provided respondents with published values of P(C) and P(+PSA), we elicited two related conditional probabilities—the PSA test's posterior probability, P(C|+PSA), and its sensitivity, P(+PSA|C)—whose ratio, but not absolute level, is constrained by the published values provided to respondents:

[5] The medical literature refers to the probability of a positive test result conditional on the presence of the disease as the test's *sensitivity* and the probability of cancer conditional on a positive test result as its *positive predictive value*. We adopt the term *sensitivity* but use *posterior probability* in place of *positive predicted value*, following custom in economics.

$$P(+PSA|C) / P(C|+PSA) = P(+PSA) / P(C) = 0.05/0.025 = 2.$$

The benefit of the elicitation technique is that respondents can know nothing about the published point estimates of the PSA test's posterior probability and sensitivity yet conform perfectly to the restriction. Percentage deviations from that restriction provide a continuous measure of an individual's inconsistency[6]:

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

Inaccuracy of an individual's beliefs with respect to published point estimates is defined as:

$$\textit{inaccuracy} = (|\log(\textit{incidence}/0.177)| + |\log(\textit{mortality}/0.028)|)/2.$$

This definition computes inaccuracy by averaging a respondent's absolute percentage deviations of lifetime incidence and lifetime mortality from their respective point estimates in the medical literature.[7]

Figure 1 shows a scatterplot of *inconsistency* and *inaccuracy*. Notice that the most inaccurate individual is among the 24 respondents who are perfectly consistent (i.e., those with zero inconsistency distributed vertically along the y-axis). The most inconsistent individual is below the midpoint of the inaccuracy range. The bivariate regression line relating inaccuracy and inconsistency has a negative but statistically insignificant slope. Based on a theory of general intelligence or a single-dimensional spectrum of axiomatic rationality as is implicit in much of the behavioral economics literature, one would expect adherence to distinct normative notions such as accuracy and consistency of beliefs to be positively correlated. These data provide no support for such a theory. Furthermore, this absence of unconditional correlation is not driven by the influence of outliers. The five most inconsistent individuals have larger-than-average levels of inaccuracy. Removing these five extreme points results in a much more negatively sloped regression line (correlation -0.31). The regression line also becomes more negatively sloped after removing the five most inaccurate observations.

Because we were also interested in the possibility of systematic bias for which the direction of deviation from published benchmarks might be important, we examined the variables *signed*

[6] We repeated the data analyses reported below with alternative definitions of inconsistency using different transformations of the percentage-point deviation, P(+PSA|C)-2 P(C|+PSA)), rather than the absolute percentage deviation above, without discovering any substantive sensitivities. We also tried simple binary categorization of the sample into Bayesians versus non-Bayesians, which we do not report, although it actually strengthened the case for our interpretations of the data.

[7] Lifetime incidence and lifetime mortality are used because the PSA frequencies (posterior probability and sensitivity) were already used to compute inconsistency. The paper's main empirical results are nearly identical using alternative definitions based on other combinations of elicited frequencies. For example, we tried an alternative definition of inaccuracy incorporating deviations from all five elicited values, which revealed no substantive changes.

inconsistency and *signed inaccuracy*, defined similarly to the definitions of inconsistency and inaccuracy above, but without taking absolute values:

signed inconsistency = log[P(+PSA|C) / P(C|+PSA)] – log[2] ,

signed inaccuracy = [log(*incidence/*0.177) + log(*mortality/*0.028)]/2.

Figure 2 shows empirical distributions for *signed inaccuracy*, *inaccuracy*, *signed inconsistency* and *inconsistency*. We discuss a few examples to aid in interpreting the units in which *inaccuracy* and *inconsistency* are measured. An individual with *signed inaccuracy = 0.10* is someone who provided mortality and incidence beliefs that were, on average, 10 percent too large. An individual with *inaccuracy = 0.10* provided mortality and incidence beliefs that were, on average, 10 percent too large *or* too small. Although log transformations help normalize the data and attenuate the influence of extreme observations on the right-hand tail, approximations of percentage deviations by log transformations become imprecise for very 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—98 percent too small rather than "390 percent too small," since -3.9 = log(0.02). The untransformed percentage deviation, [(*incidence/0.177 -1)* + (*mortality/0.028-1)*]/2, has a range of -0.98 to 9 (i.e., some respondents have beliefs that are 98 percent too small while others have beliefs that are 900 percent too large).

Accuracy and consistency within subsamples

Table 2 presents bivariate contrasts between consistent and inconsistent subsamples using four cuts of the sample by varying degrees of inconsistency: perfect Bayesians versus deviators from Bayes; below- versus above-median levels of inconsistency; bottom versus upper quartiles of inconsistency; and a fourth cut that divides the sample into those who could be modeled as Bayesians with noise versus those who commit gross errors in probabilistic reasoning (described in more detail below), referred to as Emersonians. The first column provides mean levels of *inaccuracy*, *signed inaccuracy*, log deviations for individual elicited beliefs, *inconsistency* and *signed inconsistency*. Reading horizontally across the first row, Table 2 indicates that the 24 perfect Bayesians in the sample (those with *inconsistency=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). By the fourth cut of the sample, logical inconsistency is again negatively associated with inaccuracy (1.08

among the consistent versus 0.78 among the inconsistent).

The second row of Table 2 shows that the beliefs of consistent individuals tend to be too small, whereas the beliefs of inconsistent individuals tend to overshoot published benchmarks. Consistent individuals' beliefs are not, however, generally nearer to those benchmarks. Rows 3 and 4 show log deviations for incidence and mortality, the two averaged components in *signed inaccuracy* and (with absolute values) in *inaccuracy*. Of the 16 *t* statistics in the middle block of Table 2 under the heading "log deviations," five have magnitude greater than 2, indicating statistically significant unconditional differences between consistent and inconsistent individuals. Among these five, consistent individuals' mean deviation from zero is smaller in three cases[8] and larger in two.[9] These disaggregated bivariate contrasts, while mixed, do not show any tendency for consistent individuals to have more accurate beliefs, which is generally consistent with the scatterplot of *inconsistency* and *consistency* in Figure 1.

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. Working toward a more refined taxonomy of respondents and their decision processes, we make further distinctions among those who deviate from probabilistic logic in various manners and extents. By defining three types of gross violations of probability theory, we introduce another sample cut to separate those who might be modeled as noisy Bayesians from those who, by contrast, because their beliefs cannot possibly be reconciled with the definition of conditional probability, seem to use a wholly different process to generate frequency beliefs.

A respondent is designated as *Emersonian* if any of the following three logical errors involving set intersection is present in his response profile. The residual category, *Non-Emersonians*, includes everyone else, both perfect Bayesians and those who are close enough to the consistency benchmark that one could credibly model their process as Bayesian with noise.

The first gross logical error is P(C|+PSA) > 0.50. The definition of conditional probability states that

[8] 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 Non-Emersonians and Emersonians, with *t* statistic -2.5; and -0.11 versus 0.67 for log(posterior/0.34) among non-Emersonians and Emersonians, with *t* statistic -7.9.

[9] 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.

$P(C|+PSA) = Pr(C \cap +PSA)/Pr(+PSA)$. The numerator refers to an intersection of events, for which it must be true that $Pr(C \cap +PSA) <= \min\{Pr(C), Pr(+PSA)\}=0.025$. The unconditional probabilities that were provided to all respondents imply the following logical bound:

$$P(C|+PSA) <= 0.025/0.05 = 0.50.$$

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) violate the logical bound by stating that $P(C|+PSA) > 0.50$.

The second gross error in logic that shows up in the data is $P(C|+PSA) > P(+PSA|C)$. Substituting the definition of conditional probability for both terms, the numerators of both probabilities are the same, while the denominators are known. The inequality $P(C)=0.025 < P(+PSA)=0.05$ implies

$$P(C|+PSA) <= P(+PSA|C),$$

which holds with equality only when the intersection probability in the numerator is zero. Eleven respondents violate this condition, 9 of whom also committed the first gross logical error.

A third logical error is $P(C|+PSA) = P(+PSA|C)$. 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 $P(C|+PSA) < P(+PSA|C)$. Sixteen respondents stated equal conditional probabilities. Of these, seven violated the first logical restriction by stating $(C|+PSA) = P(+PSA|C)>0.50$. Seven others stated $P(C|+PSA) = P(+PSA|C)=0.50$. Forty-five respondents committed at least one of the three errors described above, resulting in the designation *Emersonian*.

Section 3: Inconsistency predicts inaccuracy, PSA decisions?

Statistical predictors of inaccuracy?

Table 3 shows results from a regression of inaccuracy on inconsistency, with controls for information acquired, the mode in which information was processed, social influencers, age, and subfield indicators. The main finding is that, compared to the slope of the regression line for the bivariate scatterplot shown earlier, the elasticity of inaccuracy with respect to inconsistency is little changed by the inclusion of controls: -0.08 versus -0.06 with no controls. Consulting written information paradoxically appears to increase inaccuracy[10], whereas consulting written information and weighing pros and cons roughly cancel each other out. 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 overall average.

Statistical predictors of the PSA decision?

Table 4 presents estimates of four linear probability models, with *t* statistics computed using robust standard errors.[11] The first three models predict the probability of deciding to have a PSA as a function of perceived benefits, costs, and age (*fundamental* model*)*; information acquisition, information processing, and inconsistency *(add info-processing* model*)*; and social influencer indicators (*add influencers* model*).* We find statistical confirmation of the self-reported fact that most economists do not weigh costs and benefits by testing the joint hypothesis that the first five regressors have zero coefficients. The second-to-last row of Table 4 shows p-values for that hypothesis, which reveal the weak predictive power of subjective costs and benefits in the first two models. This weak predictive power does not result from overall weakness of the prediction equation, however, as likelihood ratio tests of the hypothesis that all coefficients in the model are zero are easily rejected across all models. According to the p-value in the third model, subjective costs and benefits begin to have statistically significant predictive power once social influencer variables are added. Even so, the perceived risk of incontinence has only moderate effects across the three PSA-decision models, implying that an individual whose perceived incontinence risk is twice as big as average is, at most, 6 to 8 percentage points less likely to take a PSA. Coefficients on information acquisition and processing (i.e., pros-cons deliberation and logical inconsistency) are nowhere statistically significant.

In contrast, the *doctor influenced* variable reveals strong conditional correlation between reliance on a doctor's recommendation and taking the PSA test, despite the obvious incentive problems built into the doctor-patient relationship that economists are well aware of (see Behrens, Güth, Kliemt and Levati, 2005; Loewenstein, 2005; and Sorum et al, 2004, for explanations of doctor-patient incentive

[10] One explanation for the positive association between PSA taking and consulting written information could be distorted media coverage of PSA-related medical research, giving the impression that screening is widely recommended for all men over 50, even though a number of prominent medical research organizations recommend against it.

[11] Logit and probit models produce qualitatively identical results and are available from the authors upon request. The linear probability model has the advantage of easily correcting for heteroscedasticity of errors.

mismatch).

Statistical predictors of the PSA recommendation?

The dependent variable in the fourth model presented in Table 4 is a forced-choice survey item asking for recommendations about PSA screening for an asymptomatic man in his 50s. Recall that unconditional rates of recommendation were almost twice 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 the consistency of beliefs plays at most a limited role.

Section 4: Other indications of inconsistency

As mentioned earlier, 16 individuals in the sample report that they perceive harms, yet did not weigh pros and cons. Table 5 presents a cross-tabulation of responses to the harms and pros and cons questions, including non-responses. To examine whether the joint distribution of harms and weighing pros and cons is any different among PSA takers, Table 5 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 not noticeably different.

That most economists reported not weighing pros and cons was a considerable surprise to us. This could be rationalized for individuals who perceive zero costs or zero benefits—in such cases, there are no tradeoffs to weigh. However, when it comes to PSA testing, for which the medical literature frequently recommends weighing costs and benefits, we expected most, if not all, economists to report that they had done so, especially in cases where they perceived potential harms. We worried, in fact, that this sample item would generate almost no variation. Another surprise was that most respondents reported that they did not perceive potential harms associated with PSA screening, despite the substantial medical literature documenting harms.

Table 6 demonstrates the wide spectrum of positions on PSA screening published in well-cited articles in medical journals and advocated by medical associations with influence over medical policy in the U.S. 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 explicit costs of mass screening, which have been estimated at $12 to 18 billion per year in the U.S. and predicted to increase as the over-50 population grows (U.S. Preventive Services Task Force, 2002, p. 128), one of the main points of contention regarding PSA screening concerns the benefit of early detection. Most prostate cancers grow slowly enough 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 only limited success across various stages of the disease. While there is evidence that early detection of prostate cancer reduces disease-specific mortality, there is no evidence showing benefits in terms of 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). 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).

We asked respondents the forced-choice yes/no question: "In your opinion are there potential harms associated with PSA screening?" Eleven of 133 respondents refused to say if there were harms, and one fourth of the remaining respondents answered yes. Among the 27 who reported having consulted written information about PSA testing, one third reported harms, as opposed to one fifth of those who did not consult written information, suggesting that when economists did consult information, it was absorbed to some extent by the respondents in the sample.

Guess-50 heuristic

Another issue we looked at was the number of times respondents guessed "50 percent," to see if completely uninformed priors, or use of a *guess-50 heuristic*, was linked to our primary measures of consistency and accuracy. The maximum number of times anyone in the sample guessed 50 is twice. Interestingly, those who guessed 50 more often had more accurate beliefs, with mean *inaccuracy* of 0.71 (0.01) among the 22 respondents who guessed 50 twice versus 1.02 (0.09) among those who never did. Of the 24 perfect Bayesians, two guessed 50 twice. Emersonians and Non-Emersonians guessed 50 at roughly the same rate and, as is true elsewhere in the data, inconsistency of beliefs was uncorrelated with guessing 50.

No natural frequency effect

Previous studies have documented large differences in decision making resulting from logically equivalent representations of statistical information (e.g., the framing effect in Tversky and Kahneman, 2000). 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. Therefore, we randomly alternated between two versions of the interview script that varied the way in which probabilities were presented 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. 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 *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.

Section 5: Discussion

The main objective of the survey was to collect information that would allow us to learn more about the decision-making processes that economists use, especially in high stakes decisions such as those that determine one’s health. 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 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. We also collected information about the perceived benefits and costs of PSA testing, age, professional specialization, and social influences.

We found that individuals who reported logically consistent beliefs, in the sense of conforming to definition of conditional probability, did not have more accurate beliefs. If anything, the opposite was true. Respondents who said they collected more information about PSA testing generally had more inaccurate beliefs. Confirming economists' low self-reported rate of having weighed pros and cons, we found that their beliefs about costs and benefits of PSA screening had little predictive power for PSA decisions without adding information about social influences.

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 other people, sometimes rely on a social heuristic of following doctors' advice, which could be referred to as a *white-coat heuristic*: *See a white coat, do what it says.* The social influencer indicator variables, especially the *doctor influenced* variable, add significant predictive power to single-equation models of the PSA decision. We consider further evidence in favor of this social heuristic-based explanation below (see Di Tella, Galiani and Schargrodsky, 2007, and Glaeser and Sacerdote, 2007, on the significance of social influences). Whether trusting one's doctor is effective in a normative sense is not addressed by our findings.

The paired rows of Table 7 present mean contrasts between subsamples that correspond to different hypotheses about the underlying thought process behind PSA decisions. Comparison of these differences in rates of PSA taking among paired rows shows that the largest difference is between respondents who reported that nobody influenced them and those who reported at least one influencer (36 versus 78 percent). The second pair of rows in Table 7 shows that, among those who report social influencers, there is virtually no difference in rates of PSA taking depending on having weighed pros and cons or not. According to the third pair of rows in Table 7, those who weighed pros and cons are somewhat more likely to have taken a PSA. As with all bivariate contrasts, causality is of course unclear.

One possible explanation is that those who took PSA tests contemplated pros and cons only afterward as a rationalization for what was an automatic decision following from use of a social heuristic. The fourth pair of rows in Table 7 compares pros-and-cons-weighers who do not report harms and pros-and-cons-weighers who do. 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. Perhaps those who perceive harms have a greater need to rationalize the PSA decision by reporting that it resulted from a systematic process of weighing costs and benefits, which is consistent with the fifth pair of rows in Table 7.

The sixth pair of rows in Table 7 shows that those who consulted written information were substantially more likely to take a PSA. The two rows that follow (the seventh pair in Table 7), however, suggest that, once again, causality runs in the opposite direction from that of the standard benefit-cost model, where decisions are conditioned on information, and information is unambiguously on the right-hand side. If variable levels of information acquisition led to variable rates of PSA taking in keeping with the standard model, then we would expect those who used that information (by weighing it) to have different rates of PSA taking than those who did not. Yet weighing pros and cons had almost no mediating effect among those who collected written information, consistent with the idea that choices about whether to acquire information do not significantly influence PSA decisions. Rather, it would seem that those who decided to have a PSA acquired information about its effects after the fact, as a consequence of having made that decision rather than the other way around. The last pair of rows in Table 7 shows, once again, that consistency of beliefs plays a limited role, if any, as indicated by very similar rates of PSA taking among perfect Bayesians and extreme deviators (i.e., Emersonians).

Logical consistency undoubtedly enjoys objective normative status in particular settings, for example, when taking the GRE exam. However, a growing body of theoretical models suggests that deviations from standard normative axioms in economics, paradoxically, may have beneficial effects on individual and aggregate welfare.[12] 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

[12] Inflated beliefs about the fundamental 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).

(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 criteria to particular decision-making contexts, while providing a counterexample in which standard normative benchmarks are violated yet performance is unaffected. 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, the statistical links between these distinct 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.

In the absence of solid objective criteria for saying whether a given PSA decision is correct or not, we have pursued the question of logical inconsistency's economic consequences indirectly based on the following assumption. If logical inconsistency is economically significant, it should show up in one of two ways: either in systematically different assessments of risks and benefits, or in different PSA decision outcomes holding subjective risks and benefits constant. Our data demonstrate no evidence for economic losses through either of these potential channels.

That economists exhibit large deviations from the normative decision model is not completely surprising. According to an anecdote concerning a well-known proponent of axiomatic decision theory who was trying to decide whether to take a job offer from a competing university, deviation from normative theory was deliberate and intentionally aimed at improving performance: when asked why he didn't add up probability-weighted utilities associated with each of his options and choose according to expected utility theory, the decision theorist replied: "Come on! This is serious" (Gigerenzer, 2004, p. 62).

Our first finding, that consistency does not predict accuracy, suggests that usual notions of axiomatic or consistency-based rationality are poor proxies for context-specific or adaptation-based 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, suggests that inconsistency in this domain has a small economic cost. Because inconsistency is uncorrelated with the objective accuracy of beliefs and PSA-taking behavior, it appears that inconsistency cannot explain much variance in the payoffs associated with PSA decisions.

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.

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.

Barry, Michael J, "The PSA Conundrum," Archives of Internal Medicine, 166 (2006), 38-43.

Behrens, Johann, Werner Güth, Hartmut Kliemt, and Vittoria Levati, "Games that Doctors Play Two-layered Agency Problems in a Medical System," Discussion Papers on Strategic Interaction 2006-02, Max Planck Institute of Economics, Strategic Interaction Group (2006).

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 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.

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.

Camerer, Colin F, "Do Biases in Probability Judgment Matter in Markets?: Experimental Evidence," American Economic Review, 77(5) (1987), 981-997.

Camerer, Colin F, "The Rationality of Prices and Volume in Experimental Markets," Organizational Behavior and Human Decision Processes, 51(2) (1992), 237-272.

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.

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.

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.

Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky, “The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters,” Quarterly Journal of Economics, 122(1) (2007), 209-241.

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.

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).

Glaeser, Edward L. and Bruce Sacerdote, “Aggregation Reversals and the Social Formation of Beliefs,” NBER Working Paper 13031, (2007).

Grether, D.M. and Charles R. Plott, "Economic Theory of Choice and the Preference Reversal Phenomenon," American Economic Review, 69(4) (1979), 623-638.

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).

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.

Haltiwanger, J.C. and M. Waldman, "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," Quarterly Journal of Economics, 104(3) (1989), 463-483.

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.

Harrison, Glenn W. and John A. List, “Field experiments,” Journal of Economic Literature, 42(4) (2004), 1009-1055.

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).

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.

Laibson, D. and L.Yariv, "Safety in Markets: An Impossibility Theorem for Dutch Books," Harvard University, 2004.

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, "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.

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).

Rubinstein, Ariel and Ran Spiegler, "Money Pumps in the Market," Tel Aviv University, (2006).

Savage, L., The Foundations of Statistics, (New York, NY: Wiley, 1954).

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.

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, "Friedman's Risky Methodology," University of Nottingham, (2004).

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.

Tversky, Amos, "Intransitivity of Preferences," Psychological Review, 76 (1969), 31-48.

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.

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	Valid Observations
Cash to keep?	0.12	133
Cash for 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
own PSA decision and recommendation		
Did you have a PSA?	0.46	133
Would you recommend a PSA?	0.85	133
acquired information, 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 frequencies	Mean Elicited Value	Std Dev of Mean	Valid Observations	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 | +PSA)	0.47	0.019	128	0.34
sensitivity Pr(+PSA | C)	0.72	0.018	126	0.68
incontinence probability Pr(Incontinence|Surgery)	0.30	0.020	128	.02 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 et al (2002).

Table 2: Contrasts between consistent and inconsistent subsamples

	Grand Mean	consistent 24 Perfect Bayesians	inconsistent 101 Deviators from Bayes	t stat	consistent 60 strictly below median incon-sistency	inconsistent 65 weakly above median incon-sistency	t stat	consistent 36 weakly below 25th percentile incon-sistency	inconsistent 34 weakly above 75th percentile incon-sistency	t stat	consistent 80 Non-Emersons	inconsistent 45 Emersonians	t stat
Inaccuracy measures based on absolute log deviations (and log deviations) of elicited beliefs													
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													
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 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 3: 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 4: Estimated linear probability models for the PSA decision and recommendation

	Three PSA-Decision Prediction Models:						PSA-	
	fundamental		*add info-processing*		*add influencers*		Recommendatio	
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 5: 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 6: 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 Institute	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 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

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***□

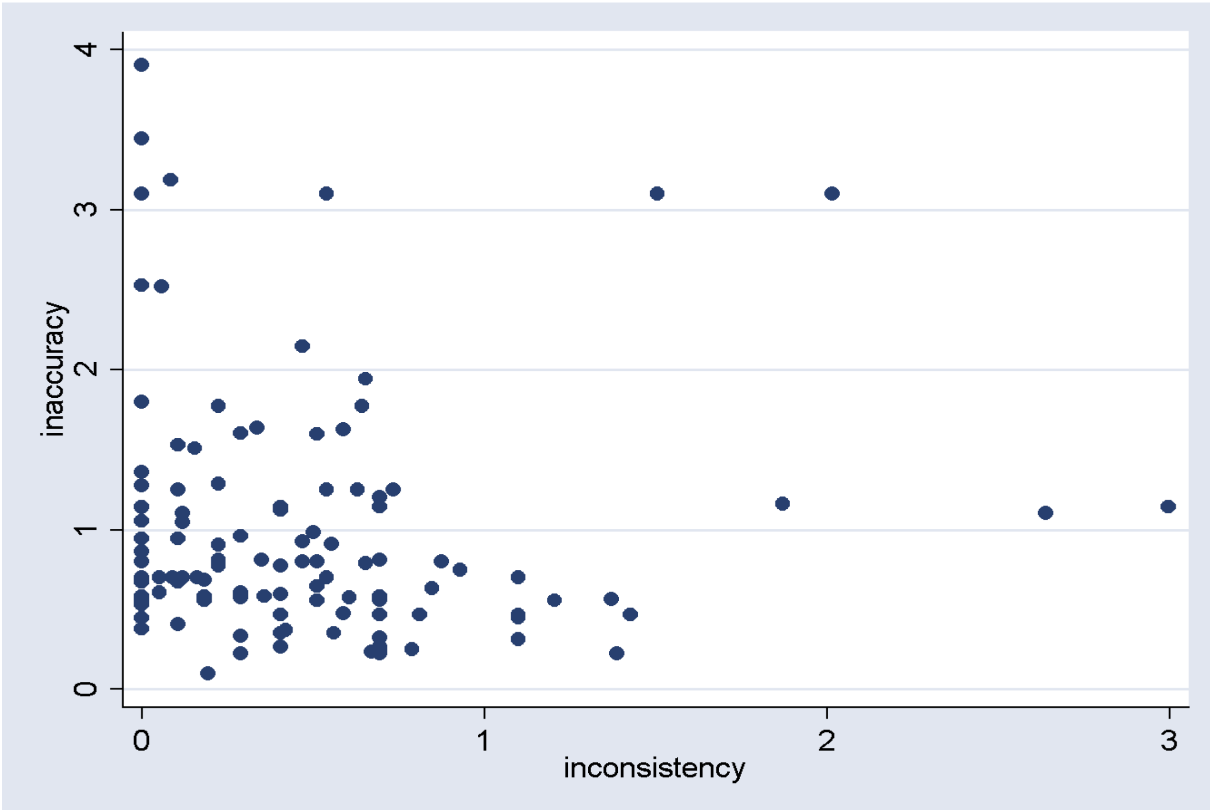
The main focus of the survey is prostate cancer and PSA (Prostate Specific Antigen) screening. I won't ask any personal questions about the illness itself, just about screening. I'd like to elicit your best guesses about the risks of prostate cancer.

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***□

Figure 1: 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.

Figure 2: Empirical distributions of unsigned inaccuacy, unsigned inconsistency, inaccuracy and inconsistency