An Old Measure of Decision-making Quality Sheds New Light on Paternalism[*]

Shachar Kariv and Dan Silverman[†]

October 27, 2012

Abstract

Definitive judgment about the quality of decision making is made difficult by twin problems of measurement and identification. A measure of decision-making quality is hard to formalize, to quantify, and to make practical for use in a variety of choice environments; and it is difficult to distinguish differences in decision-making quality from unobserved differences in preferences, information, beliefs, or constraints. In this paper, we describe a widely applicable set of tools for theoretical analysis and experimental methods for addressing these problems. These tools and methods can indicate a more targeted approach to "light paternalism" polices aimed at improving decision-making quality.

JEL Classification Numbers: D01, D02, D03, D04, P43, P46.
Keywords: Decision-making quality, rationality, revealed preference, libertarian paternalism, asymmetric paternalism, soft paternalism.

[†]Kariv: University of California–Berkeley (email: kariv@berkeley.edu); Silverman: Arizona State University and NBER (email: dsilver3@asu.edu).

1 Introduction

Behavioral economics has challenged the empirical validity of standard economic models as it has posed new questions about individual behavior and introduced new methods (for economists) of empirical research. This branch of economics research has raised important questions about the rationality of individual behavior and has inspired a variety of theoretical alternatives to standard models. The investigation of these alternative theories has uncovered still more empirical regularities that are hard to reconcile with existing models, and so on. In this way, behavioral economics has improved our understanding of many important social and economic phenomena.

An important part of the behavioral economics literature can be viewed as a response to traditional economic analysis which assumes that individuals make the choices that are in their own best interest. In this standard view, heterogeneity in the choices that people make is attributed to heterogeneity in these individuals' constraints, preferences, information, or beliefs. A prominent strand of behavioral economics adds importantly to this list of the sources of variation in behavior; that research considers heterogeneity in choices driven also by differences in the *quality* of decision making.

The idea that decisions vary in quality and individuals differ in their decision-making abilities seems uncontroversial. Almost everything else comes in different levels of quality, so we would expect some choices to be better than others and some individuals to be better decision-makers than others. When making certain economic choices, some individuals may tend to misunderstand the relevant tradeoffs; they may compute payoffs incorrectly, pay too little attention and make unintended choices, follow the behavior of uninformed others, or err in other less obvious ways. As a result, the choices that some individuals actually make may sometimes be different from the choices they would make if they had more time, skills, or knowledge, or were in a better state of mind to consider their decisions. Importantly, this view allows that, even if they have all relevant information, some individuals might not identify and make the choice which best meets their objectives.

Rigorous evidence for the view that decision-making quality is a source of heterogeneity in behavior derives mostly from studies that collect uncommonly high quality data, or exploit instrumental variables, and show that some decisions clearly leave "money on the table." Ameriks et al. (2003) is a prominent example. That paper provides evidence that differences in individuals' propensity to plan and budgeting behaviors, rather than more standard sources of heterogeneity, explain important variation in wealth accumulation. Restricting attention only to financial decision making, in another

example, Bernheim and Garrett (2003) find evidence that employer-based financial education increases saving. The literature also includes, among others, Duflo and Saez (2003) who investigate the effect of financial education on saving, beyond its effect on lifetime earnings; Lusardi and Mitchell (2007) document very low levels of financial planning, financial literacy, and a positive correlation between literacy, financial planning and wealth; and Cole and Shastry (2009) emphasize the importance of education, cognitive ability and financial literacy on financial market participation.

Motivated by evidence that individuals do not always make choices that are in their best interest, some behavioral economists have helped design "*light* paternalism" polices aimed at steering people toward better choices. We emphasize the term "light" to underscore that, in contrast to hard paternalism (such as compulsory automobile seatbelts or motorcycle helmets), the paternalism that these behavioral economists advocate is intended to avoid direct coercion. The prominent forms of light or soft paternalism are libertarian paternalism (Thaler and Sunstein, 2003) and asymmetric paternalism (Camerer et al., 2003). As emphasized by Loewenstein and Haisley (2008), "the common goal of these paternalistic approaches is to steer human behavior in more beneficial directions while minimizing coercion, maintaining individual autonomy, and maximizing choice to the greatest extent possible."

To illustrate, consider the work on light paternalism devoted to problems of retirement planning and economic choices late in life. Many analysts have long been concerned about the quality of economic choices in and near retirement, and the recent proliferation of retirement saving and health insurance plans has only heightened that concern. When pensions were mostly defined benefit and health insurance consisted of employer-provided plans and Medicare parts A and B, planning for retirement was relatively simple. Now that most pensions are defined contribution, the typical worker needs to determine a contribution rate and make a complex portfolio choice for their 401(k). Later, the worker will need to determine how best to draw down that pension wealth, and then which of the dozens of Medicare Part D plans is best. Already, research has shown low levels of basic financial knowledge and a decline in that knowledge and in more basic cognitive functioning with age. This lack of knowledge and its decline raise obvious concerns about the quality of important financial decisions like saving for retirement.[1]

---

[1] Agarwal et al. (2009) show a U-shaped age pattern in the frequency of dominated choices regarding the use of credit, with both relatively young and old consumers more prone to error. Banks (2010) summarizes the research on the relationships between cog-

In response to these concerns, some advocates of light paternalism have focused special attention on retirement saving plan design. Many employers match their employees' contributions to tax-advantaged retirement saving plans like a 401(k). And yet many employees do not take full advantage of the match and thus appear to be turning down free money. This is especially true when those employees can borrow against their retirement savings at low interest rates. Many observers, understandably, describe such choices as errors. To address these sorts of errors, employers, government agencies, and others have worked to educate employees about the nature of the plans and the employer match, and simplify the presentation of the choices. This is not paternalism, but rather a sensible effort at financial education.

The paternalism emerges in the setting of *default* saving rates in employer-sponsored saving plans. A substantial and rigorous body of research shows that individual retirement saving plan decisions are importantly influenced by the setting of saving contribution choices. Madrian et al. (2001) and Choi et al. (2004), among others, show that the setting of defaults has significant effects on the amount of saving that employees do in their employer-sponsored plans. As Thaler and Sunstein (2003) argue, default saving rates are unavoidable; they must be set at some level. And if employees are both influenced by defaults, and missing out on substantial employer matches, why not set the default saving rates higher? If, upon careful consideration, an employee views the default saving rate as too high, he can easily adjust his choice as he fills out his benefits form. If, however, an employee finds the saving decision too confusing, or too time-consuming to properly consider, why not help him exploit the employer match and tax-advantaged saving by setting a higher default?

This view of employer-sponsored saving plans has been widely adopted and has had an important influence on policy.[2] One reason for this influence may be the clarity of the relevant incentives and constraints in this domain. In this special setting, regarding some choices as lower quality is straightforward and relatively uncontroversial. More generally, paternalistic policy-makers need to adopt a concept of decision-making quality and then

---

nitive function, financial literacy and financial outcomes at older ages.

[2]In perhaps the most significant example, the Thrift Savings Plan Enhancement Act of 2009 calls for all new Federal employees in the U.S. to be automatically enrolled in the Thrift Saving Plan program (TSP) at a default saving rate of 3%, unless they opt out. The TSP is the rough equivalent of a 401(k) for employees of the U.S. Federal Government. Most recently, Senator Daniel Akaka (D-HI) has introduced the Save More Tomorrow Act of 2012 that would have the default rate be 3% in the first year of employent and rise by increments of 1% for at least the next two years.

use that concept to make judgements about what are better decisions. A central challenge for the light paternalism project is that, in many situations, a measure of decision-making quality is difficult to formalize, quantify, and to make practical and portable for use in different choice environments (the measurement problem).

As important, in most economic settings, it is unclear to the policy-maker whether people are making lower quality decisions as opposed to facing different but unobserved incentives, or having different preferences over the same outcomes, or possessing different information, or holding different beliefs. Hence, comparing the relative importance of heterogeneity in decision-making quality and heterogeneity in preferences, information, beliefs, or constraints for predicting economic decisions is particularly important (the identification problem). This comparison matters because decision-making quality, unlike preferences, may be justifiably manipulated. If differences in decision-making quality are important sources of the heterogeneity in economic outcomes, then even quite costly policy changes aimed light paternalism may hold substantial promise.

Loewenstein and Haisley (2008) suggest an "imperfect but pragmatic" solution to the measurement and identification problems. In particular, they advocate the 'cautious' approach that "even light paternalistic policies should only be put into play when welfare judgments tend to be relatively straight-forward." In their view, this is the case when individuals take strictly dominated actions, or make decisions that lead to clearly negative outcomes, or make choices are not in line with their professed preferences. Their examples include, but are not limited to, contributing less than the 401(k) match threshold, filing for Chapter 13 bankruptcy, and wanting to get out of debt but failing to do so.

We propose a different approach. First, we argue for measuring decision-making quality by the compliance of choices with economic rationality. Second, we suggest presenting individuals with simple and stylized economic choice experiments in which analysts can measure their decision-making quality with a high degree of precision, and separate it from other sources of heterogeneity in choice. Then, we propose examining whether this theoretically grounded measure of decision-making quality – the consistency of the experimental data with the utility maximization hypothesis – is useful in explaining heterogeneity in choices made in complex, real-world environments. Such choices include consumption decisions, insurance purchase, retirement decisions, and health investments among other important behaviors. The idea is that, if decision-making quality is a trait, better decision making in the experiment should predict better decision making in the real world. If

so, then experimental measures of decision-making quality can be used to design and target policies aimed at improving real-world economic choices.

The rest of this paper is organized as follows. Section 2 provides more background on the need for new data and analysis of decision-making quality. Section 3 describes our contribution to method. Section 4 discusses the implications of our approach for light paternalism. Section 5 concludes by relating the paper to the literature and contains other concluding remarks. An Appendix illustrates the merits of our approach using a prototype example, and contains some technical digressions that the interested reader may wish to consult.

2　The need for new data and analysis

Observational studies of market data, laboratory and field experiments, and surveys have all been used to measure aspects of decision-making quality. As implemented, each of these methods has advantages, but also limitations. The principal challenge that all these methods face is the difficulty of distinguishing variation in decision-making quality from the other unobserved sources of heterogeneity in choices, namely differences in preferences, information, beliefs or constraints.

Market data. Observational studies of market data have the advantage of analyzing natural decisions, often with high stakes but these observational studies are subject to identification problems (Manski, 1995): the impossibility of observing all the relevant variables limits the reliability of these analyses. Administrative datasets that draw from electronic transaction records hold important promise since are about as accurate as possible and thus provide a high resolution picture of individual choice. However, these datasets typically provide information about the activity in just one form of saving, or on just a few credit cards, or at just one retailer.

Hence, a key disadvantage of these data is that they represent just a slice of the economic activities of the individuals involved. An uncommonly high-quality administrative dataset is needed to make relevant incentives and constraints sufficiently clear, so that regarding some decisions as "mistakes" is straightforward and uncontroversial. To this end, the data must give more than just a small piece of the individual's financial picture. Only a comprehensive dataset can reveal whether a lower quality choice represents a minor bobble in otherwise sound decision making, or a more fundamental problem in evaluating economic choices.

Laboratory experiments. Experiments can overcome many of the identification problems of observational study. But the typical experiment has elicited relatively few decisions from a relatively small and homogenous sample. Also, while these experiments reveal that "mistakes" exist, they give us little sense of how important they are or how frequently they occur. Because the samples are small and homogenous, the results are also rarely linked to important behaviors outside the laboratory so external validity is a concern (Levitt and List, 2007).

A variation of these criticisms is that the experimental treatments typically used are not very important relative to choice scenarios in "the wild." Field experiments draw on large and heterogenous samples, and link experimental and "natural" behaviors. The obvious limitation relative to laboratory experiment is that, although we may be able to control some variables, there are many important variables that we cannot control in the field.

Survey. Large-scale surveys of consumption, income, wealth and portfolio typically rely on the self-reports of individual households. Accurate responses to survey questions about financial matters require considerable effort from participants, effort that may be greater for those with fewer economic resources or cognitive abilities. Inaccuracy in the reporting muddles the picture on decision-making quality, raising the question of whether a choice was actually low quality, or simply mis-reported.

An obvious response to the relative limitations of these methods is to combine them and take advantage of each of their strengths. Analysts could integrate survey research with administrative records and experiments to help distinguish the role of decision-making quality from other sources of heterogeneity in choice. For example, large-scale surveys could include experimental interventions aimed at eliciting, in simple and stylized settings, direct information about preferences such as risk tolerance, time discounting, or elasticities of intertemporal substitution. This information could then be used to determine the extent to which heterogeneity in stable preferences underlies heterogeneity in behavior. Indeed, the Health and Retirement Study (HRS) has been pioneering such integrated methods for the past 20 years, implementing survey questions designed to isolate often abstract, but powerful concepts from economic theory. Several large panel surveys, conducted all over the world, have since followed suit.

Until very recently, however, the integration of theory-based experiment and survey research has been limited to hypothetical questions with limited scope. A highly influential example is the risk tolerance questions pioneered

by the HRS. These questions ask:

> "Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50-50 chance the second job would double your total lifetime income and a 50-50 chance that it would cut it by a third. Which job would you take – the first job or the second job?"

Subsequent questions unfold to separate respondents into as many as five different levels of risk tolerance. These questions are superbly tailored to capture a conceptually central notion, and the responses have been used widely in economics research. The questions are, however, relatively limited in the amount of information they can reveal about individuals' risk attitudes (heterogeneity is limited to the five categories of risk tolerance). In addition, because they pertain to a hypothetical situation with which almost no one has any experience, they demand a great deal of imagination from respondents and offer no direct incentive to exercise that imagination.

Laboratory experiments usually include monetary incentives; but technical limitations have both restricted the power of incentivized experiments to reveal individual preferences and kept them in university labs with student subjects. Typically, the experiments involve a limited number of decisions. Subjects are often presented with binary choices and the decisions often involve extreme choices such as choices between lotteries in which some outcomes have very small probabilities. Indeed, many such experiments are designed only to discover violations of standard theories. In addition, until recently, it was very difficult to link incentivized, experimental choices to the behavior of large and heterogenous samples outside the laboratory. The traditional, pencil-and-paper, or telephone methods of large surveys made it extremely costly to conduct such experiments.

To summarize, for those interested in studying economic decision-making quality, there are important obstacles to obtaining reliable solutions to the measurement and identification problems. The combination of survey and experimental methods holds promise in this regard, but only recently have technological constraints been lifted, making a more complete integration possible. In the next section we outline and argue for some specific efforts in this vein. More specifically, we argue for an approach that combines revealed preference analysis of experimental data with economic and demographic information from a survey. In what follows, we describe our preferred criterion

for decision-making quality and discuss a "tool kit" that allows for (1) convincing identification of individual behaviors in the experiment, and (2) relating individual-level behaviors from the experimental data with economic information and socio-demographic information on individuals.

3   An old criterion and a new tool kit

3.1   The criterion

In his Foundations of Economic Analysis (1947), Paul Samuelson offered a criterion for measuring decision-making quality based solely on observable behavior. Adopting Samuelson's approach, we suggest calling choices lower quality (only) if there is no well-defined (utility) function that the choices maximize. Classical revealed preference theory provides a direct test of this notion of quality: choices are consistent with maximizing a utility function if and only if they satisfy the Generalized Axiom of Revealed Preference (GARP). In our view, if there is no utility function that choices maximize then those choices cannot be considered purposeful and, in this way, high quality.

We should emphasize that Samuelson's criterion for decision-making quality is not as restrictive as it sounds. It 'only' requires consistent preferences over all possible alternatives, and choices that correspond to the most preferred alternative in the feasible set. Any consistent preference ordering is admissible. Furthermore, since GARP imposes the complete set of conditions implied by utility-maximization, goodness-of-fit indices provide a stringent test of decision-making quality. We refer the interested reader to the Appendix for more details on the tests for consistency with GARP.

We also want to emphasize that, in our view, consistency with GARP is a necessary condition for decisions to be considered high quality, but it is not sufficient. Choices can be consistent with GARP and yet fail to be reconciled with any utility function that is normatively appealing given the decision problem at hand. For example, a natural question to ask about data on choice under risk is whether choices are consistent with the *dominance principle* in the sense of Hadar and Russell (1969) – that is, the requirement that an allocation should be preferred to another, regardless of subjects' risk attitudes, if it yields unambiguously higher monetary payoff. The dominance principle is compelling and generally accepted in decision theory. Clearly, violations of monotonicity with respect to first-order stochastic dominance should be regarded as errors.

3.2 The tool kit

3.2.1 The field environment

We think that studies which combine laboratory methods and survey research hold important, and still largely untapped, promise. Web-based surveys like the CentERdata and the American Life Panel (ALP) offer a new opportunity to improve on existing methods.

- The CentERdata is a research institute in the Tilburg School of Economics and Management (TiSEM) in the Netherlands. CentERdata specializes in online surveys and manages several panels, each consisting of more than 2,000 households, that are representative of the Dutch-speaking population in the Netherlands. Via the Internet, CentERdata (www.centerdata.nl) can implement sophisticated experiments and collects a great deal of individual demographic and economic information from its panels.

- The ALP (www.mmicdata.rand.org) is a 4,000 member, U.S.-based, Internet panel administered by the RAND Corporation. Like the CentERdata panels, it can implement sophisticated experiments and collects extensive demographic and economic information from its members. Distinctively, the ALP collects information, using survey methods developed for the HRS, on topics central to economic theory. CentERdata and ALP provides an uncommon opportunity to combine experimental and survey data.

3.2.2 The laboratory environment

Choi et al. (2007a, 2007b) (hereafter, CFGK) introduced an experimental technique that allows for the collection of richer data about preferences than has previously been possible and can be adapted to a wide range of decision-making experiments in large-scale surveys. In the experiment, CFGK present subjects with a standard economic decision problem that can be interpreted either as a portfolio choice problem (the allocation of wealth between two risky assets) or a consumer decision problem (the selection of a bundle of contingent commodities from a standard budget set). These decision problems are presented using a novel graphical experimental interface.[3] With this interface, subjects see on a computer screen a geometrical

---

[3] Ahn et al. (2011) extend the experimental work of CFGK in settings with risk (known probabilities) to settings with ambiguity (unknown probabilities). Fisman et al. (2007) employ a similar methodology to study social preferences.

representation of the budget set and choose allocations through a simple "point-and-click."

As usual, the experiments avoid the problems of identification and interpretation that hamper observational study and unincentivized survey response. In addition, and more than typical laboratory experiments, the experimental allows the rapid collection of large amounts of choice data and thus permits a stringent test of utility maximization, and tight bounds on preferences. We refer the interested reader to the Appendix for more details on the experimental design.

In ongoing research, we and our colleagues implement versions of the experiment in CFGK using the CentERdata and can provide: (a) regression analysis of the relationship between preferences, the degree of rationality, and socio-demographic variables, and (b) investigations of the correspondence between decision-making quality in an experiment and economic outcomes in the real world. In Choi et al. (2012) we take up the question of whether consistency with revealed preference as measured in the experiment can help explain heterogeneity in the wealth holdings of households with similar lifetime incomes. Using different surveys and populations, the same methods could be used to investigate whether heterogeneity in this measure of decision-making quality can help explain heterogeneity in consumption choices, insurance purchases, retirement decisions, or health investments among other important behaviors.

3.3   The methodological contribution

We believe that the entire apparatus – analytical and experimental techniques – has a number of useful features for purposes of evaluating economic decision-making quality:

- **Portable.** The analytical techniques and experimental platforms are applicable to *any* type of individual choice problems involving personal and social consumption. They can thus make domain-specific predictions and provide a *unified* measure of decision-making quality across domains.

- **Autonomous.** Consistency with utility maximization is not affected by the underlying preference type and the experimental task makes no special demands of outside knowledge or expertise, thus helping to isolate heterogeneity in decision-making quality from heterogeneity in preferences, information, beliefs or constraints (the identification problem).

11

- **Quantifiable.** A variety of goodness-of-fit indices measure the *extent* of GARP violation. Hence, in contrast with hypothetical (and unincentivized) survey questions, we can understand the results in terms of economic theory, which helps interpret (as well as design) the experiments.

We view the revealed preference approach as complementing those which use tests of cognitive ability (IQ) to measure decision-making quality and capacity (cf. Dohmen et al., 2010). The measures of compliance with GARP complement IQ scores by offering a metric for evaluating the quality of decisions that is firmly grounded in economic theory. More specifically, measures of compliance with GARP have well-established economic interpretations that permit, among other things, the disentangling of quality from preferences. In addition, classical theory tells us whether we have enough data to make tests of compliance with GARP statistically useful. There is no comparable, theoretically disciplined, means of implementing, interpreting, and evaluating an IQ test.

Another distinction of the GARP test relative to IQ tests is that the former is easily portable to a variety of choice problems. We can thus make domain-specific predictions and study a comparable measure of decision-making quality across domains. In addition, consistency with GARP avoids the need for right and wrong answers, and the task can be designed to make very limited demands of outside knowledge or expertise. Virtually all IQ tests have right and wrong answers, and thus draw on outside knowledge and depend on preferences for obtaining certain skills. For example, Raven's matrix tests, spatial relations tests, and number series tests, all have right and wrong answers and all involve skills developed by training in Mathematics.

4 Implications for light paternalism

Much more must be done to evaluate the robustness and usefulness of our proposed approach to studying decision-making quality. We need to apply the experimental platforms and analytical techniques that have been developed to other types of individual choice problems, and to further assess the intrapersonal persistence of the results. If, however, the methods prove robust we argue they can be used to refine and target efforts at light paternalism.

One over-simplified, but useful, view of the experiments is that they can distinguish two groups of individuals. The first group of individuals can –

at least in a controlled experimental setting where the tradeoffs are sufficiently transparent – make rational economic decisions. Call this group of individuals "Grade A" decision-makers. Individual in this group are capable of making purposeful choices in the sense that these choices achieve a well-defined objective. The second group of individuals, by contrast, do not make rational economic decisions, even in an experimental setting. Call this group "Grade B."

In our view, Grade A decision-makers merit greater deference from policymakers; the usual caution that is applied in implementing light paternalism policies should be even higher for them, and more resources could be wisely allocated to educating them about the choices that they face. The reason to exercise greater caution is that we suspect that most Grade A decision-makers will, in complex real-world settings, face many of the same distractions and challenges in making choices as the rest of us, and will thus be similarly influenced by paternalistic interventions. We therefore suspect that, despite their Grade A status, they too will respond to even a light touch of paternalism. But Grade A decision-makers can meet their various objectives when the setting is right and the tradeoffs are is transparent enough. Respect for their autonomy thus indicates that, for this group, light paternalistic policies should take a backseat to efforts at clarifying the tradeoffs they face (like balancing risk and return or consumption and saving) and the mapping between their choice sets and their objectives.

Why not show the same deference to Grade B decision-makers and exhaust all clarifying efforts before turning to light paternalism? Clarifying tradeoffs can be expensive; it often requires substantial amounts of financial education and can be quite specific to the setting. Grade B decision-makers show that, even in a relatively simple and controlled setting, they cannot make rational decisions. The return to efforts at clarifying the tradeoffs they face in the real-world is therefore likely to be very low. When resources are scarce, light paternalism aimed at steering individuals Grade B decision-makers toward sensible choices is thus more appropriate.

5   Concluding remarks

In standard, neoclassical economic models, decision-makers choose their preferred alternative from the feasible set given the information available to them, and heterogeneity in choices is attributed to heterogeneity in preferences, information, beliefs, or constraints. Allowing that some decisions may not be rational, a strand of behavioral economics considers heterogeneity in

choices driven also by differences in the quality of decision making. This literature takes seriously the idea that individual choices are often subject to error, and that these errors are systematic. This research thus allows some of the choices that some people actually make to be different from the choices they would make if they had the skills, insights, time, and state of mind to make "better" decisions.

If people systematically fail to make choices that are in their best interest, it is logical for economics research to work toward alleviating those failures. Thus the turn to design light paternalism policy aimed at steering people toward choices that promote their own welfare. While light paternalism holds considerable promise and has already had an important effect on policy-making, the idea remains a controversial subject of debate. Many analysts, including some leading proponents of light paternalism, see important challenges to the design and implementation of policies aimed at improving economic decisions. In this paper, we offered two contributions to that debate. First, we provide a simple and clear statement of central problems that attempts at light paternalism must face. Second, we describe a widely applicable set of methods for addressing those problems.

More specifically, we argue that attempts at light paternalism must face the central problems of measurement and identification. An answer to the measurement problem is a definition conceptualizing decision-making quality in a formal way so it gives us a precise sense of how good individuals are in making economic decisions. Given a definition of decision-making quality, an answer to the identification problem is a widely applicable method for separately identifying decision-making quality from standard mechanisms that generate variation in choice.

- As an answer to the measurement problem, we propose compliance with utility maximization as a necessary condition for high quality decision making. In order to be considered high quality, that is, choices must be coherent and thus reconcilable with a utility function. In this sense, we take the view high-quality decisions are purposeful choices; we insist only that they pursue some objective.

- As an answer to the identification problem, we propose now well-tested experimental methods and econometrics that, with limited assumptions and based on choice (price and quantity) data alone, can distinguish differences in decision-making quality from (typically unobserved) differences in preferences, information, beliefs, or constraints.

In on-going work with Syngjoo Choi and Wieland Müller (e.g., Choi

et al., 2012), we are providing proof of concept, and demonstrating the potential of these answers to the measurement and identification problems by combining experimental methods with web-based survey research.

6 Appendix

6.1 The CFGK design and procedures

In the CFGK experiment, subjects make decisions under conditions of uncertainty about the objective parameters of the environment. In the preferred interpretation, there are two *states of nature* denoted by $s = 1, 2$ and two associated *Arrow securities*, each of which promises a dollar payoff in one state and nothing in the other. CFGK consider the problem of allocating an individual's wealth between the two securities. Let x_s denote the demand for the security that pays off in state s and let p_s denote its price. The budget set is then $p_1 x_1 + p_2 x_2 = 1$ and the individual can choose any portfolio $(x_1, x_2) \geq 0$ that satisfies this constraint.

In the experiment, each decision problem started with the computer selecting a budget line randomly from the set of budget lines that intersect with at least one of the axes at 50 or more tokens, but with no intercept exceeding 100 tokens. Each choice involved choosing a point on a budget line of possible token allocations. To choose an allocation, subjects used the mouse to move the pointer on the computer screen to the desired allocation.

During the course of the experiment, subjects were not provided with any information about the account that had been selected in each round. At the end of the experiment, the computer selected one decision round for each subject, where each round had an equal probability of being chosen, and the subject was paid the amount he had earned in that round. Payoffs were calculated in terms of tokens and then converted into dollars.

6.2 Testing for consistency with GARP

Following Afriat's (1967) theorem, CFGK employs GARP to test whether the finite set of observed price and quantity data that our experiment generated may be rationalized by a utility function.[4] The broad range of budget

---

[4] Varian (1982, 1983) provide precise details on testing for consistency with GARP. The papers by Afriat (2012), Diewert (2012), Varian (2012) and Vermeulen (2012), published in a special volume of the *Economic Journal* on the Foundations of Revealed Preference provide an excellent overview and a discussion of some recent developments in the literature.

sets faced by each subject provides a rigorous test of GARP. In particular, the changes in endowments and relative prices are such that budget lines cross frequently. This means that the data lead to high power tests of revealed preference conditions. To assess how nearly the data complies with GARP, CFGK calculate Afriat's (1972) Critical Cost Efficiency Index (CCEI).

The CCEI measures the fraction by which each budget constraint must be shifted in order to remove all violations of GARP. By definition, the CCEI is between 0 and 1: indices closer to 1 mean the data are closer to perfect consistency with GARP and hence to perfect consistency with utility maximization. Figure 1 illustrates the construction of the CCEI with a simple violation of the Weak Axiom of Revealed Preference (WARP) involving two allocations, x^1 and x^2. It is clear that x^1 is revealed preferred to x^2 because $p^1 \cdot x^1 > p^1 \cdot x^2$, yet x^1 is cheaper than x^2 at the prices at which x^2 is purchased, $p^2 \cdot x^1 < p^2 \cdot x^2$. If we shift the budget constraints by $A/B < C/D$ as shown, then the violation would be removed.

[Figure 1 here]

Note that although the CCEI provides a summary statistic of the overall consistency of the data with GARP, it does not give any information about which of the observations are causing violations. Alternative measures have been suggested by Varian (1990, 1991) and Houtman and Maks (1985). Varian (1990, 1991) refined Afriat's CCEI to provide a measure that reflects the minimum adjustment required to eliminate the violations of GARP associated with each budget constraint. Houtman and Maks (1985) finds the largest subset of choices that is consistent with GARP. Using these disaggregated measures, the revealed preference framework can be used to construct a choice-based welfare criterion to evaluate paternalistic interventions by limiting welfare calculations to the subset of consistent choices, as advocated by Bernheim and Rangel (2008) and further developed by Bernheim and Rangel (2009). We refer the interested reader to Loewenstein and Haisley (2008) for a discussion of the different welfare criteria proposed in the literature.

References

[1] Afriat, S. (1967) "The Construction of a Utility Function from Expenditure Data." *Econometrica, 6*, pp. 67-77.

[2] Afriat, S. (1972) "Efficiency Estimates of Production Functions." *International Economic Review, 8*, pp. 568-598.

[3] Afriat, S. (2012) "Afriat's Theorem and the Index Number Problem." *The Economic Journal, 122*, pp. 295-304.

[4] Agarwal, S., J. Driscoll, X. Gabaix and D. Laibson (2009) "The Age of Reason: Financial Decisions over the Life-Cycle with Implications for Regulation." *Brookings Papers on Economic Activity, 2*, pp. 51-117.

[5] Ahn D., S. Choi, D. Gale and S. Kariv (2012), "Estimating Ambiguity Aversion in a Portfolio Choice Experiment." Revise and resubmit *Quantitative Economics.*

[6] Ameriks, J., A. Caplin and J. Leahy (2003) "Wealth Accumulation and the Propensity to Plan." *Quarterly Journal of Economics, 118*, pp. 1007-1047.

[7] Banks, J. (2010) "Cognitive Function, Financial Literacy and Financial Outcomes at Older Ages: Introduction." *Economic Journal, 120*, pp. 357-362.

[8] Bernheim, D. and D. Garrett (2003) "The Effects of Financial Education in the Workplace: Evidence from a Survey of Households." *Journal of Public Economics, 87*, pp. 1487-1519.

[9] Bernheim, D. and A. Rangel (2008) "Choice-Theoretic Foundations for Behavioral Welfare Economics." In *The Foundations of Positive and Normative Economics*, ed. A. Caplin and A. Schotter. Oxford University Press.

[10] Bernheim, D. and A. Rangel (2009) "Beyond Revealed Preference: Theoretic Foundations for Behavioral Economics." *Quarterly Journal of Economics, 124*, pp. 51-104.

[11] Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003) "Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism." *University of Pennsylvania Law Review, 151*, pp. 1211-1254.

[12] Choi, J., D. Laibson and B. Madrian (2004) "Plan Design and 401(k) Savings Outcomes." *National Tax Journal, 57*, pp. 275-298.

[13] Choi S., R. Fisman, D. Gale and S. Kariv (2007a) "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit," *American Economic Review Papers & Proceedings*, *97*, pp. 153-158.

[14] Choi S., R. Fisman, D. Gale and S. Kariv (2007b) "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, *97*, pp. 1921-1938.

[15] Choi S., S. Kariv, W. Müller and D. Silverman (2011) "Who Is (More) Rational?" Revise and resubmit *American Economic Review*.

[16] Cole, S. and G. Shastry (2009) "Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation." Harvard Business School Finance Working Paper No. 09-071.

[17] Diewert, W (2012) "Afriat's Theorem and some Extensions to Choice under Uncertainty." *The Economic Journal*, *122*, pp. 305-331.

[18] Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010) "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, *100*, pp. 1238-1260.

[19] Duflo, E. and E. Saez (2003) "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence From a Randomized Experiment." *Quarterly Journal of Economics*, *118*, pp. 815-842.

[20] Fisman R., S. Kariv and D. Markovits, (2007) "Individual Preferences for Giving." *American Economic Review*, *97*, pp. 1858-1876.

[21] Hadar, J. and W. Russell (1969) "Rules for Ordering Uncertain Prospects." *American Economic Review*, *59*, pp. 25-34.

[22] Houtman, M. and J. Maks (1985) "Determining all Maximial Data Subsets Consistent with Revealed Preference." *Kwantitatieve Methoden*, *19*, pp. 89-104.

[23] Levitt, S. and J. List (2007) "What do Laboratory Experiments Tell Us About the Real World?" *Journal of Economic Perspectives*, *21*, pp. 153-174.

[24] Loewenstein, G. and E. Haisley (2008) "The Economist as Therapist: Methodological Ramifications of "Light" Paternalism." In *The Foundations of Positive and Normative Economics*, ed. A. Caplin and A. Schotter. Oxford University Press.

[25] Lusardi, A. and O. Mitchell (2007) "Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth." *Journal of Monetary Economics, 54*, pp. 205-224.

[26] Madrian, B. and D. Shea (2001) "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics, 116*, pp. 1149-1525.

[27] Manski, C. (1995) "Identifcation Problems in the Social Sciences." Cambridge: Harvard University Press.

[28] Samuelson, P. (1947) *Foundations of Economic Analysis.* Cambridge: Harvard University Press.

[29] Smith, J., J. McArdle and R. Willis (2010) "Financial Decision Making and Cognition in a Family Context." *Economic Journal, 120*, pp. 363-380.

[30] Thaler, R. and S. Benartzi (2004) "Save More Tomorrow: Using Behavioral Economics to Increase Employee Savings." J*ournal of Political Economy, 112*, pp. 164-187.

[31] Thaler, R. and C. Sunstein (2003) "Libertarian Paternalism." *American Economic Review*, Papers & Proceedings, *93*, pp. 175-179.

[32] Varian, H. (1982) "The Nonparametric Approach to Demand Analysis." *Econometrica, 50*, pp. 945-972.

[33] Varian, H. (1983) "Non-Parametric Tests of Consumer Behaviour." *Review of Economic Studies, 50*, pp. 99-110.

[34] Varian, H. (1990) "Goodness-of-Fit in Optimizing Models." *Journal of Econometrics, 46*, pp. 125-140.

[35] Varian, H. (1991) "Goodness-of-Fit for Revealed Preference Tests." Mimeo.

[36] Varian, H. (2012) "Revealed Preference and its Applications." *The Economic Journal, 122*, pp. 332-338.

[37] Vermeulen, F. (2012) "Foundations of Revealed Preference: Introduction." *The Economic Journal, 122*, pp. 287-294.

Figure 1. The construction of the CCEI for a simple violation of WARP