

GARP for Kids: On the Development of Rational Choice Behavior

By WILLIAM T. HARBAUGH, KATE KRAUSE, AND TIMOTHY R. BERRY*

Do children choose rationally? This question matters for several reasons. If children are incapable of choosing rationally, then there is little point in attempting to use standard economic models to describe their behavior. For example, models of family bargaining behavior assume that family members each act rationally to maximize a utility function. On a more applied level, families and society engage in a wide variety of paternalistic policies towards children. These policies use various mechanisms to encourage certain choices and discourage or prohibit others. The issue of rational choice by children is important both for justifying this paternalism and for determining the effectiveness of incentive-based mechanisms for enforcing it. It is also difficult to justify using data on children's choices to draw inferences about their preferences, or to accurately predict their future behavior, if the choices are not rational.

In addition to these reasons, which apply to children as children, there is the fact that children grow into adults. An understanding of the development of rational choice behavior over the lifespan may ultimately tell us something about rational choice by adults. Is the ability to choose rationally something that is universally present at very young ages, or does it only appear at adulthood, or even later? Is rational choice the rule, or the exception, among children? If rational choice is not the rule, then do

children who reason better also choose more rationally?

In this paper we report on the results of an experiment that tests whether children make rational choices about consumption goods. We studied 7- and 11-year-old children and, for comparison, college undergraduates. The experiment tests variations on what might be seen as the most basic requirement for rationality, namely that choices must obey transitivity. If a person picks A when given a choice between A than B, and B when given a choice between B and C, then barring indifference rationality requires that he must pick A when given a choice between A and C. We also examine how rationality, as measured by several different tests of transitivity and by a simple measure of the size of the violations, changes with age and mathematical ability.

Section I explains the relevant aspects of the theory that relates choice behavior and utility maximization in more detail. This is followed by a discussion of some previous experiments on adults and animals in Section II, and then by Sections III and IV discussing our protocol and results, respectively; Section V concludes.

I. Theory

Revealed preference theory began with Paul A. Samuelson (1938), who showed that it was possible to think about rational choice purely in terms of observable data, without recourse to unobservable constructs like utility. Hendrik S. Houthakker (1950) and Sidney N. Afriat (1967) gave necessary and sufficient conditions for choice data to be consistent with utility maximization. Hal R. Varian (1982) refined the main theorem in Afriat's paper, and proved that satisfying the Generalized Axiom of Revealed Preference (GARP) is a necessary and sufficient condition for choice data to be consistent with the maximization of a continuous, concave, locally nonsatiated, and weakly monotonic utility function. This means that if choices satisfy

* Harbaugh: Department of Economics, University of Oregon, Eugene, Oregon 97403, and National Bureau of Economic Research (e-mail: bill@harbaugh.org); Krause: Department of Economics, 1915 Roma NE, Economics Building, University of New Mexico, Albuquerque, NM 87131 (e-mail: kkrause@unm.edu); Berry: Department of Economics, University of Oregon, Eugene, Oregon 97403 (e-mail: trb@gladstone.uoregon.edu). We thank Anne van den Nouweland, Jim Andreoni, an anonymous referee, and participants in the 1999 NBER Children's Program meeting for helpful comments. This research was funded by National Science Foundation Grant Nos. SBR-9810835 and SBR-9810847.

GARP we lose nothing by using standard economic models to analyze behavior. This leads naturally to the idea of testing the otherwise unobservable hypothesis of utility maximization, by checking to see if some set of observed choice data is consistent with GARP.¹ In this section of the paper we explain what this test entails.

Prior theoretical work assumes that choices are made from continuous budget sets, defined by prices and incomes. Since many of our participants have trouble doing the math needed to stay within a budget constraint, in our experiment we had them make choices from finite sets of allowable bundles. This means we need to restate GARP in terms of choices, as follows. First, we say that a person directly reveals that he prefers bundle x^i to bundle x when he chooses x^i over x or over a bundle with at least as much of every good as in x , and more of at least one. We say that a person indirectly reveals that he prefers x^i to bundle x when some sequence of directly preferred relations between bundles connects x^i to x . GARP then requires that if a person directly or indirectly reveals that he prefers x^i to x , he cannot choose bundle x^i when some alternative x^j with at least as much of every good as in x , and more of at least one, is available. We can then state a new version of Afriat's theorem: If choices satisfy this version of GARP, they are consistent with the maximization of a continuous, concave, strongly monotonic utility function.

Figure 1 shows why we need to strengthen the original requirements of weak monotonicity and local nonsatiation to strong monotonicity. Suppose a and b are chosen from the choice sets A and B respectively, where each choice set consists of the bundles indicated by the dots that are connected by the respective line. By weak monotonicity we can conclude $u(a) \geq u(b)$, since when a is chosen something at least as good as $b = (2,3)$, namely $d = (2,4)$, was in

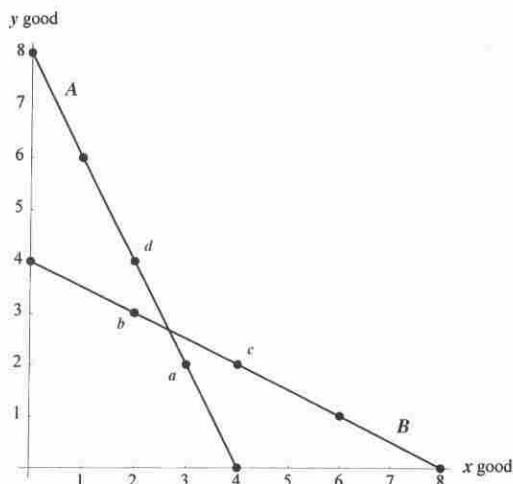


FIGURE 1. REVEALED PREFERENCE WITH DISCRETE CHOICE SETS

the choice set. However, this is not enough to reject rational choice, since the only other thing we know is that $u(b) \geq u(a)$, by similar logic, and if $u(a) = u(b)$ then neither of these choices is irrational. Local nonsatiation does not add any bite here, since it only requires that there be some alternative within any given distance of b that provides higher utility: it does not require that the alternative actually be in the choice set.

However, if we assume utility functions are strongly monotonic, we can use revealed preference to test rationality. Now we can use these choices to say $u(a) > u(b)$, since a was chosen over d , which by strong monotonicity must provide more utility than b . A similar argument would imply $u(b) > u(a)$, since b was chosen over $c = (4,2)$ which must provide more utility than $a = (3,2)$. The contradiction means that these choices cannot be the result of rational choice.

In addition to counting violations, for consistency with other papers we report a version of Afriat's (1972) efficiency index, a measure of the severity of the GARP violations.² This measure was developed in the context of budget sets, and is based on the fact that a choice that

¹ Samuelson's result used the Weak Axiom of Revealed Preference, which only makes direct comparisons between choices and does not allow indifference curves to have straight segments. Houthakker used the Strong Axiom of Revealed Preference, which allows indirect comparisons but maintains the indifference curve assumption. GARP allows for both indirect comparisons and straight segments of indifference curves. The conclusions in this paper are the same using any of these definitions of rational choice.

² We count violations using what we see as the simplest method. For each of the choices we count one violation as occurring if that choice, in combination with any of the other choices, violates GARP. An alternative would be to count each such violation separately.

violates revealed preference can be interpreted as a waste of money. When a revealed preference violation occurs, the person could have made some alternative choice that he revealed he preferred, instead of the choice he actually made. The cheaper this alternative is, the more money was wasted by not choosing rationally. The index e measures the overall efficiency of the participants' choices, or $1 - e$ measures the proportion of income that a person wasted by making the choice that violated revealed preference. An index of one means either no violations or that an infinitesimally small change in a choice would eliminate all violations. A discussion of the advantages and disadvantages of this index is available in Varian (1991).

To calculate this index with our choice data we use the implicit prices and incomes from which our choice sets are derived, and we make the simplifying assumption that the choice sets are continuous. Revealed preference violations and the efficiency index were obtained by using an algorithm that was modified from Varian (1995), and which is available from the authors.

II. Previous Experiments

Economists have done many studies to test whether observed choice data can be reconciled with the axioms of revealed preference. Some studies have looked at changes in household consumption in response to market price and incomes fluctuations. Such empirical studies tend to have little power to reject rational choice, because the observed price fluctuations are so small and because the data used typically lumps goods into categories. A classic early experimental alternative is Raymond C. Battalio et al. (1973), which used data on consumption choices by women patients at a psychiatric hospital. The participants bought goods at a commissary, and the authors arranged to have the prices of goods periodically changed by large amounts, and records kept of individual purchases. Depending on the way in which errors in the recording of choice decisions are accounted for, they found that between 5 and 50 percent of the participants made choices that violated revealed preference.

One problem with this study, and similar studies that use either controlled or market-induced price fluctuations, is that if preferences change over time, choices made at different

times cannot be used to test for rational choice. Three more recent papers have used a random lottery procedure to avoid this problem. This procedure involves giving the participant a list of intersecting budget sets, only one of which will be actually implemented. The participant's problem is to choose so as to maximize expected utility. So long as preferences are independent of irrelevant alternatives, this procedure elicits the data necessary for checking consistency with the revealed preference axioms, namely the participant's most preferred alternative from a variety of different sets of alternatives.

James Andreoni and John Miller (2002) used this procedure to look at 142 college students' decisions about how much money to keep for themselves and how much to share with another under eight different budget constraints. They found that 9 percent of the participants had some violations of the revealed preference axioms. Harbaugh and Krause (2000) replicated this experiment on children and found considerably more violations. Reinhard Sippel (1997) studied 42 college students' choices for eight different consumption goods, using ten different budget sets. He found that 24, or more than half, of the participants violated GARP.

Developmental psychologists have also done work on intransitivity in children, but with a very different focus. The typical experiment is designed as a test of reasoning, not of choice, and involves either questions like "If Ted is older than Bill, and Bill is older than Jim, then who is older, Ted or Jim?" or similar comparisons of the length or weight of objects. The general conclusion of the literature, as reviewed in Leonard Breslow (1981), for example, is that while it is difficult to measure this ability while controlling for the development of general verbal skills, it appears to develop between the ages of 7 and 9. It seems quite possible that a child could perform poorly at transitive reasoning, yet be perfectly able to avoid intransitive choices.

There is also a related literature on choice by animals, focusing on whether choices by pigeons and rats obey the laws of demand, summarized in John H. Kagel et al. (1995). Typically, in these experiments food is allocated by pushes on levers, and incomes and prices are varied by limiting the number of allowed lever pushes and varying the amount of

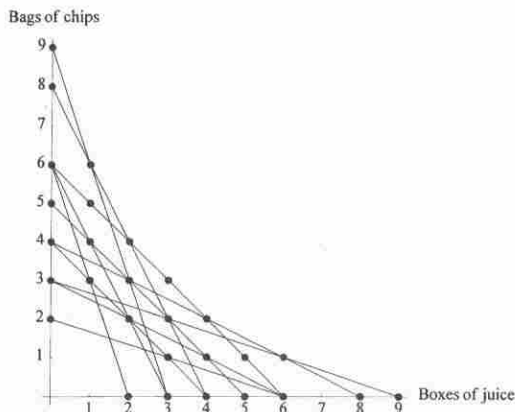


FIGURE 2. CHOICE SETS

food dispensed per push. The participants are given one budget constraint for about two weeks at a time, by which time consumption has stabilized. Then the constraint is changed. A typical experiment would involve testing for an inverse relationship between price and quantity by imposing a new budget constraint passing through the original consumption bundle, but with different relative prices. They find 87 percent of the rats and 81 percent of the pigeons reduce consumption of a good when given an income-compensated price increase. Note that this test of rational behavior is much less rigorous than those done on humans.

III. Protocol and Participants

Our experiment is similar in design to those of Andreoni and Miller and of Sippel. We presented our participants with 11 different choice sets. Each choice set was a list of between three and seven bundles, with each bundle consisting of a number of small bags of potato chips and a number of boxes of fruit juice. We used goods that would typically be consumed quickly because we wanted as little possible interaction between decisions in the experiment and outside influences. Figure 2 shows the choice sets as dots representing the bundles included in each choice set. As can be seen by the lines connecting the dots, the choice sets we used consisted of all the integer combinations of bundles on the frontier of a budget set that was determined by a fixed income and prices.

The participants were given a stapled packet,

with each page listing the bundles for a single choice set. We told them they would be asked to circle the bundle that they liked best from each page. After they were done choosing we would pick one page, and they would get to actually keep the bundle they picked from that page. Sufficient supplies of chips and juice were shown to make this promise credible. We told them not to mark their choice until 30 seconds were up, and to spend any extra time looking over the bundles and making sure they had really picked the one they liked best. It was clear that 30 seconds was more than enough time to allow for careful decisions.

After we had gone through all 11 choice sets, we repeated the procedure, this time giving them 15 seconds for each choice set. We told them to think some more about the bundles, and make any changes they wanted to. We then picked one choice set, showed it and their choice to each participant, and gave them a third and final chance to change their chosen bundle. They were then given that bundle in a paper sack. The script and details of the experiment are in the unpublished Appendix (available upon request from the authors).

This protocol provides the sort of data necessary for testing rational choice, namely multiple observations of individual choices from intersecting choice sets. Since there is a probability that every choice will be implemented, the participants have an incentive for making choices that actually represent their preferences.

The experiments were done on 31 second-grade students and 42 sixth graders in Eugene, Oregon public schools, and also on 55 college undergraduates at the University of Oregon. We did not record individual ages, but second graders are generally about 7 years old, sixth graders about 11, and these undergraduates average about 21. The grade-school experiments were arranged with the permission of cooperative teachers, and all students in their class that day participated. Since school attendance is mandatory, and the proportion of children in private and home schools is small, this method of recruiting participants ensures that our sample of children is very representative of the local population.

The undergraduates were included to provide a plausible upper limit, for this protocol, against which the amount of rationality in children's choice performance could be compared. We

first tried using performance by graduate students and faculty in economics as our standard. However, these participants all had some familiarity with revealed preference theory, and in postexperiment interviews they stated that they were primarily concerned with avoiding what they saw as embarrassing violations of rationality, and only secondarily with getting the bundles of goods that they preferred. This preference clearly affected their behavior—most picked corner solutions, explaining later that they knew this would ensure their choices would satisfy GARP. (Those who did not adopt this strategy did exhibit some violations of revealed preference.) To provide a more realistic baseline, we ran the comparison experiment on undergraduate economics majors. This was done at the first meeting of an undergraduate intermediate microeconomics class and before any mention of revealed preference theory.

The undergraduate experiments were run with the same protocol as were the experiments on children. We should note that college undergraduates are not only older than the children, but are also a more select group in terms of mathematical ability and general intelligence. On the other hand, they are probably less motivated by the actual consumption of these particular goods. Still, they were clearly interested, and were willing to wait in line for as long as five minutes after the completion of the experiment to get their chips and juice. We believe that their performance is close to the maximum degree of rationality that we should reasonably expect to find in adults using this protocol.

IV. Results and Discussion

We use the results for the participants' final choices.³ In the protocol, the bundles from each choice set were always presented to the partici-

³ We also looked at whether changes in choices led to fewer violations. The second graders made an average of 2.0 changes, sixth graders 1.2, and undergraduates 0.75. For each grade we regressed the final number of violations against the initial number of violations and the number of choices that were changed. For all grades the coefficient on the initial number of violations was positive and significant, while that on the changes was negative and significant for the second graders and actually positive, but insignificant, for the older kids and undergraduates. We interpret this to mean that the older participants who change their choices do not have strong preferences over these goods to begin with. We think the close similarity of these results for sixth graders and undergraduates strengthens the conclusion that these groups have similar choice behavior.

TABLE 1—AVERAGE REVEALED PREFERENCE VIOLATIONS, BY AGE

Group	Average number of GARP violations	Afriat's Index
Second graders	4.3	0.93
Sixth graders	2.1**	0.96
Undergraduates	2.0	0.94
Random (uniform)	8.91	0.648
Random (bootstrapping)	8.29	0.749

Notes: The mean numbers of violations and the mean of Afriat's Index for all the age-groups are less than for random choice at the 1-percent significance level, using *t*-tests. For the age-groups, ** indicates significantly different from the mean directly above at the 1-percent level.

pants in descending order by the number of juice boxes. We used a consistent order so that participants could quickly grasp the possible consumption choices without having to search through the bundles. To check that the order in which the choice sets were presented did not affect violations, we showed the choice sets in four different orders. We found that orderings did not have significant effects on the severity index or on the number of violations of any of the tests of revealed preference (nor on preferences for one good over another).

Table 1 shows average violations of GARP for the three different age-groups. The last two lines of the table gives results for two different versions of random choice, each with 10,000 observations, where each observation consists of one choice from each of the 11 different budget sets. For the first version we pick each observation's choice from a given choice set randomly and with equal probability from the set of possible choices in that set. The second is a bootstrap-type simulation. For each of these observations the choices for a given budget set are drawn using weights reflecting the proportion of our sample making that choice from that budget set. This gives a measure of the number of violations that would be expected from random choice, while incorporating information about people's actual choices, rather than just the possible alternatives.

Four results are apparent from this table. Some revealed preference violations are present at all ages. Even the youngest children have considerably fewer and less severe violations than would be expected if they were choosing randomly. The number of violations decreases noticeably from age 7 to 11. From age 11 to 21, there is only a

TABLE 2—PERCENTAGE OF GROUP WITH DIFFERENT NUMBERS OF GARP VIOLATIONS

Number of GARP violations	Percentage of group with violations			Random (bootstrapping)
	Second graders	Sixth graders	Undergraduates	
0	26	62	65	2
1	0	0	0	0
2	3	7	2	5
3	10	5	9	4
4	6	0	0	9
5	26	7	13	11
6	0	7	2	15
7	6	2	0	17
8	10	5	0	17
9	10	2	4	12
10	3	2	2	7
11	0	0	4	2

small and insignificant further decrease in the rationality of choices.

We then ask whether the violations are driven by a few people making many violations, or by small numbers of violations by many people. Table 2 shows the distribution of the number of GARP violations, again across the age-groups. The distribution from the random choices is included for comparison.

The distribution of violations in our participants is clearly very different than that from random choice. Even for second graders (the group with the most violations), 26 percent of the participants have no violations, compared to 2 percent under bootstrapping. For this particular test of rationality, we can say that about 25 percent of the 7-year-olds and about 60 percent of 11-year-old participants make choices that are consistent with utility maximization. We can also say that for this test there is no real increase in rational choice from age 11 to 21.

While the temptation is difficult to resist, we cannot use this table to conclude anything like "only 26 percent of second graders can choose rationally" or "11-year-olds are more than twice as likely to choose rationally as are 7-year-olds." Our experiment is only one of a variety of possible tests, and participants who pass this test might well fail a more rigorous one. Similarly, we cannot conclude that, in general, rational choice does not substantially increase from the age of 11 to 21. We might well expect a larger increase with a more rigorous test, though it is also conceivable that we would find a smaller one.

The test by Sippel is more difficult than ours in that it involves choices from nearly continu-

ous budget sets over ten goods rather than just two, but he also gives the participants more time to decide (he used 20–30 minutes, versus 9 or so in our protocol). Given these countervailing factors, his result that 57 percent of the participants violate GARP seems generally consistent with what we find. Andreoni and Miller's study of altruistic choices finds GARP violations in only 9 percent of the participants. Their participants face 8 budget sets rather than the 11 in our experiment, and perhaps preferences over sharing are better defined even than those over familiar consumption goods.

Last we address the question of whether children with better mathematical ability are better at choosing. Our measure of mathematical ability is performance on the multiple-choice part of the Oregon Mathematics Problem Solving Assessment Test. This is an hour-long test that was designed to measure student achievement and was administered to the sixth graders when they were in fifth grade. We had data for 37 of the 42 sixth-grade participants. The mean score was 227 and the standard deviation was 11. Table 3 shows the results of regressions of the number and severity of GARP violations in the last round against performance on this test score. Nearly identical results were found for the first round. In both regressions the coefficient had the "correct" sign, in the sense that higher mathematical ability was correlated with more rational choices. While the magnitude is large, particularly for the number of GARP violations, where a two-standard-deviation increase in the score is associated with a 25-percent decrease in the number of violations,

TABLE 3—RATIONALITY VIOLATIONS AND MATHEMATICAL ABILITY IN SIXTH GRADERS

	Average number of GARP violations	Afriat's Index for GARP
Math score	-0.0323 (0.0462)	0.000940 (0.00179)
R^2 (37 observations)	0.014	0.0078

neither effect is statistically significant. We also found that performance on this test was not significantly correlated with the number of choices that were changed between the first and last rounds of the experiment. In addition, students with higher scores were not more likely to decrease their number of violations, either in general or conditional on the number of changes.

V. Conclusion

The argument that people behave rationally is central to economics, and the experiment reported in this paper is perhaps the simplest possible test of economic rationality. It requires no ability to forecast choices by others and act strategically, and no ability to think about time or probability in any rigorous way. We just ask people to choose what bundle of goods they like from a list of a few alternative possibilities—in our protocol, they do not even need to do the math needed to stay within a budget constraint.

Using this experiment, we find that at age 7 children's choices about consumption goods show clear evidence of rationality, though also many inconsistencies. By age 11, choices by children with below-average mathematical ability are as rational as choices by adults with above-average intelligence, although even these adults' choices show many inconsistencies. Based on our results, we conclude that, to the extent the assumption of utility maximization is useful for modeling choice behavior by adults, it is also appropriate for children. We are not claiming that children's behavior in more complicated economic situations, for example those involving choices over time or under uncertainty, or requiring strategic behavior, can be described using the same models as are used for adults. These issues are subjects for future research.

REFERENCES

- Afriat, Sidney N.** "The Construction of Utility Functions from Expenditure Data." *International Economic Review*, February 1967, 8(1), pp. 67–77.
- _____. "Efficiency Estimation of Production Functions." *International Economic Review*, October 1972, 13(3), pp. 568–98.
- Andreoni, James and Miller, John.** "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 2002 (forthcoming).
- Battalio, Raymond C.; Kagel, John H.; Winkler, Robin C.; Fisher, Edwin B.; Basmann, Robert L. and Krasner, Leonard.** "A Test of Consumer Demand Theory Using Observations of Individual Consumer Purchases." *Western Economic Journal*, December 1973, 11(4), pp. 411–28.
- Breslow, Leonard.** "Reevaluation of the Literature on the Development of Transitive Inferences." *Psychological Bulletin*, March 1981, 89(2), pp. 325–51.
- Harbaugh, William T. and Krause, Kate.** "Children's Altruism in Public Good and Dictator Experiments." *Economic Inquiry*, January 2000, 38(1), pp. 95–109.
- Houthakker, Hendrik S.** "Revealed Preference and the Utility Function." *Economica*, May 1950, 17(66), pp. 159–74.
- Kagel, John H.; Battalio, Raymond C. and Green, Leonard.** *Economic choice theory: An experimental analysis of animal behavior*. Cambridge: Cambridge University Press, 1995.
- Samuelson, Paul A.** "A Note on the Pure Theory of Consumer's Behavior." *Economica*, February 1938, 5(17), pp. 61–71.
- Sippel, Reinhard.** "An Experiment on the Pure Theory of Consumer's Behaviour." *Economic Journal*, September 1997, 107(444), pp. 1431–44.
- Varian, Hal R.** "The Nonparametric Approach to Demand Analysis." *Econometrica*, July 1982, 50(4), pp. 945–73.
- _____. "Goodness of Fit for Revealed Preference Tests." Working Paper No. 13, University of Michigan, September 1991.
- _____. "Efficiency in Production and Consumption." Working paper and Mathematica notebook, 1995. [Online:] <http://emlab.berkeley.edu/eml/nsf97/varian.pdf> [January 24, 2000].