The family under the microscope: an experiment testing economic models of household choice.

Alistair Munro and Ian J. Bateman and Tara McNally

National Graduate Institute for Policy Studies, Japan

2. June 2008

The family under the microscope: an experiment testing economic models of household choice.

by

Alistair Munro,

Ian J. Bateman

and

Tara McNally

Alistair Munro (Corresponding author. Email to alistair-munro@grips.ac.jp): National Graduate Institute for Policy Studies, Tokyo, Japan and Department of Economics, Royal Holloway College, University of London, Egham, TW20 0EX, **Ian J. Bateman**: School of Environmental Sciences and CSERGE, University of East Anglia, Norwich NR4 7TJ, UK. **Tara McNally**: Department of Economics, Royal Holloway, University of London, Egham, TW20 0EX, UK. The work reported here was financed by the UK Economic and Social Research Council, award number RES-000-22-2081.

Abstract.

We devise and execute three experiments to test key features of models of household decision-making. Using established couples (married and unmarried) we test income pooling, unanimity and Pareto efficiency. Subjects make choices individually and jointly and are asked to make predictions about their partner's choices. Unanimity is rejected. Income pooling is not rejected in joint choice but has less explanatory power in individual choice. In direct tests both sexes do not pool income completely, but in econometric tests across all tasks, women place an equal weight on payoffs but men discount their partner's payoffs by between 15 and 20%. We find that transparency has little impact on deviations from income pooling or indeed on behaviour generally. Many joint choices deviate from the Pareto principle in a systematic manner suggesting that choices made as a couple are more risk averse than individual decisions.

JEL Codes: C920, D130, D80.

Keywords: experiment, household, unitary, income pooling, Pareto, family.

Introduction.

There is a large body of experiments on *individual* choice, but to date there is little formal experimental work that tests economic models of *household* behaviour. Within psychology there is a tradition of research that examines descriptive models of family decision-making, usually on the basis of data obtained from expenditure diaries, questionnaires and stated preference methods (see Corfman and Lehman, 1987). However, the relationship between the results of this literature and the formal models of economists is at best unclear. One of the few economics experiments on household behaviour is an enterprising study by Elizabeth Peters et al, 2004. It employs a variety of family groups in a public goods game. Compared to the contributions made within groups composed of strangers it finds that contributions made within family groups are typically higher and the reductions in contributions over time are much weaker. A second study, by Bateman and Munro, 2005, finds that households do not conform to the predictions of expected utility theory. Ashraf, 2005, meanwhile examines how control and transparency affects household savings decisions for couples in the Philippines and finds that male behaviour is more sensitive to context compared to that of women. None of these papers are focused on direct tests of key aspects of household models such as efficiency and whether members pool income – the main concerns of much of the theoretical and empirical work on the family. This paper therefore presents experiments designed to test some of the main predictions of competing economic models of the household. To achieve this we recruit established couples and face them, individually and jointly with a sequence of binary choices which involve payoffs for themselves and for their partners.

One difficulty of testing economic theories of the household is that they differ not only in the equilibrium notion involved, but also in the level of detail given about the process which leads to the proposed outcome. Should a theory, for instance, apply only to the choices made by the household collectively, or should it apply equally to the choices made by individuals within that household? [1] Given the ambiguity over exact predictions, we concentrate on testing some key properties that are features of a number of common variants of the main models. By comparing individual and collective choice we test the Pareto principle. By manipulating which individual receives the payoffs while maintaining constant aggregate household payoffs we test the principle of income pooling (i.e. the property that, given prices, patterns of choice should be independent of who receives the income in the household). By comparing individual and joint choices we test for unanimity within the household. By varying the privacy of decisions we test whether asymmetric information affects behaviour.

There are of course a large number of tests of household models using market data (e.g. Lundberg et al, 1997, Phipps and Brown, 1998, and Browning and Chiappori, 1998 or see Doss, 1996 and Alderman et al, 1995 for surveys) which in many cases yield apparently conflicting results. When published consumption data from developed nations is used (e.g. Browning and Chiappori, 1998) the unitary model of Becker, 1974, is typically rejected in favour of the Paretian alternative. In other contexts the situation is less clear: for instance Brolin *et al* 2002 reject a Paretian interpretation of household investments in child health. Meanwhile, Jones, 1983, rejects

---

[1] Consider a household which makes collective choices according to $\mu(m_1,m_2)u^1(x_1) + (1-\mu(m_1,m_2))u^2(x_2)$, where u^i, i=1,2 is the riskless utility function for the ith individual, m_i is income received by the ith individual, x_i is the consumption of the ith individual and μ represents the weight attached to individual 1 in the collective decision. Suppose individual 1 must make a choice between (10, 2), (8, 12) and (2,20) where (a,b) means £a paid to the individual and £b to their partner. They might choose according to $\mu u^1 + (1-\mu)u^2$ (the *collective* individual) or they might choose according to $u^1(m_1)$ (*myopic*) or they could choose according to u^1 knowing that any eventual allocation of resources will be according to $\mu u^1 + (1-\mu)u^2$ (*Stackleberg).* In the second case they will choose (10,2), but in either of the other cases they may choose one of the other two options, depending on μ and its relation to the values of m_i. If, say, $u^i=\log(x^i)$, and $\mu = (m^1/(1+m^1+m^2))^{0.5}$, then the collective choice is (2,20) and the Stackleberg choice will be (8,12).

intra-household efficiency in the Cameroon, as does Udry, 1996 for farm level data in Burkina Faso, while Udry and Duflo, 2001, also reject Pareto efficiency for farm households in the Côte d'Ivoire. These differences in results may be the result of different economic contexts or may reflect data quality, but either way they suggest the value of complementary evidence provided by controlled tests of the predictions of household models.

In our data, we find little support for unanimity. We find that the choices made jointly are often in conformity with income pooling, but that when separated partners frequently make different choices and often these do not conform to income pooling. Instead individuals – especially men - tend to overweight their own income compared to their partners. The degree of privacy of the decisions has little or no impact. Perhaps most surprisingly, we find that the Pareto principle is often breached, typically because in joint choice, couples are more risk averse than when they face the equivalent choice separately.

The remainder of the paper is as follows: properties to be tested are presented formally next in section II; Section III covers the detailed design of the experiment, while the results are presented in section IV. Section V concludes the paper.

II Properties of Household Models.

Consider a two-person household (or an n person household where only the preferences of two agents matter directly for decisions). Index agents by i=1,2 and states of the world by s=1,…,S. Depict a lottery, L, where agent i receives payment m_{is} in state of the world s, by the vector $(m_{11},\ldots,m_{1S},m_{21},\ldots,m_{2n})$ and let $L^i = (m_{i1},\ldots,m_{iS})$, i=1,2. We say two lotteries, L^A, L^B are Income Pooling Equivalent (IPE) if, for all s=1,…,S, $m_{1s}^A + m_{2s}^A = m_{1s}^B + m_{2s}^B$. Let $\succeq$ be the weak preference relationship for the household (with strict preference, $\succ$, and indifference, $\sim$, constructed in the usual manner) and let $\succeq_i$ represent weak preference for individual i = 1,2.[2]

We now state four potential properties of preferences. The final two are relationships between individual and household preference. The first two properties are stated for the household; with minor modifications of notation they can also hold at the level of the individual.

Income Pooling (IP). If L^A, L^B are IPE then for all L^C, $L^A \succeq L^C \leftrightarrow L^B \succeq L^C$.

Interpersonal Separability (S). For all L^A, L^B, L^C, L^D such that for one i, $L^{Ai} = L^{Bi}$ and $L^{Ci} = L^{Di}$ and for j≠i, $L^{Cj} = L^{Aj}$ and $L^{Dj} = L^{Bj}$, $L^A \succeq L^B \Rightarrow L^C \succeq L^D$.

Pareto Dominance (P). $L^A \succeq_1 L^B$ and $L^A \succeq_2 L^B \Rightarrow L^A \succeq L^B$. $L^A \succ_1 L^B$ and $L^A \succeq_2 L^B$ or $L^A \succeq_1 L^B$ and $L^A \succ_2 L^B \Rightarrow L^A \succ L^B$

Unanimity (U). $L^A \succeq_1 L^B \Leftrightarrow L^A \succeq_2 L^B \Leftrightarrow L^A \succeq L^B$. $L^A \succ_1 L^B \Leftrightarrow L^A \succ_2 L^B \Leftrightarrow L^A \succ L^B$

IP is usually associated with the unitary model, but IP is also an implication of the non-cooperative model (Chen and Woolley, 2001) if both agents make strictly positive contributions to the household public goods and such contributions are made ex post - i.e. after the state of the world is known, (Warr, 1981). IP can also be compatible with the Paretian model if the weight attached to individual utilities is independent of income shares.[3] We shall call individuals *income poolers* if their choices satisfy IP. We refer to lottery A as *IP-safer* (riskier) than B, when from an income pooler's perspective, A is safer (riskier) than B in the sense of Rothschild and Stiglitz, 1970. A lottery is IP safe when it has no risk when viewed by an income pooler.

The second property, separability, implies that the household or individual will be

---

[2] Lotteries allow greater generality in our approach. Additionally, they help to divert attention away from the between subjects trade-offs implicit in the experiment and thus may reduce experimenter induced other-regarding behaviour (List and Levitt, 2005).

[3] Typically in Paretian models the weights are not independent of income shares. However, for small changes in income shares we might expect them to be approximately constant.

indifferent about the correlation between L^1 and L^2. If for instance, the household maximizes $\mu E(u^1) + (1-\mu)E(u^2)$, where u^i , i=1,2 is utility, μ is the weight attached to person 1's utility and E is the expectations operator. If ex post transfers are not feasible, and μ is insensitive to changes in the pattern of household income then S will hold. The property P is a key prediction of all Paretian models. It implies that if two individuals both prefer A over B, then acting jointly as a household, they will not prefer B to A. Finally, unanimity is self-explanatory.

None of the properties make the assumption that either individual preferences or the preferences of the household satisfy the assumptions of expected utility theory (EUT). EUT has a long history of experimental scrutiny, much of which is hostile to the assumptions of the model and as we have noted Bateman and Munro, 2005 rejects EUT for households. As a result we design our experiment without assuming EUT.

The properties are the main hypotheses to be tested, but there are subsidiary conjectures that are also worth examining or particular directions for the alternative hypotheses. Opportunities for partners to hide or disguise some of their expenditure and income abound within many households (Pahl, 1990). Theoretical models of the household are usually written in the context of full information, but it would be useful to know if individuals choose differently when their choices are hidden from their partners. An obvious though by no means compelling conjecture in this case is that individuals will behave more selfishly when choices are private and certainly Ashraf, 2005, finds this amongst men in her experiment. Additionally, if there is blame attached to making the wrong decision, there may be an incentive to choose more riskily when choices are hidden and opportunities for recrimination are reduced. Similarly when we think about how behaviour may deviate from income pooling, the most obvious alternative is that individuals place greater relative weight on their own payoffs. It is also conceivable that deviation occurs because individuals weight risk differently if it appears in their partner's payoffs. Table 1 summarises the tests, along with alternative hypotheses where these are one sided.

Table 1. Main Hypothesis.

Null hypothesis	Alternative hypothesis and comments.
1. Individual behaviour is unaffected by privacy.	1. Individuals are more selfish when choices are private. 2. Individuals are more risk loving when choices are private.
2. Unanimity – individual and joint choices concur.	
3. Joint choices respect income pooling.	
4. Individual choices respect income pooling	1. Individuals place more weight on their own payoffs, compared to their partners. 2. Individuals place more weight on risks to their own income, compared to risk in their partners' income.
5. Choices respect interpersonal separability.	Subjects prefer payoff combinations that are negatively correlated between partners.
6. Joint choices are Pareto efficient.	

III Experimental Design.

We report data from three linked experiments, A, B and C that share the same basic structure of a three-part design, summarised below. Before getting into details it is worth outlining the differences between the experiments. As the name suggests, experiment A was the prototype. In early runs it became clear that the experiment could accommodate more questions in the self-imposed one hour time slot[4]. Moreover, some minor hypotheses were quickly rejected by the accumulating data. We replaced some tasks and added some, producing experiment B with 36

[4] Conscious of the time constraints on our subjects, most of whom were in paid employment and/or had children, we designed the experiment to fit into a one hour slot.

questions (n=12) rather than 30 (n=10). Twenty-four tasks were common to both datasets. Experiment C is a follow-up to B with the same number of questions. The fundamental difference between B and C is in the information available to each partner about their other partner's choices and payoffs. In A and B we took great care to keep this information private (see below); in C we made all actions and payoffs visible between partners. Not all the questions were the same in B and C, but the first group were common to both experiments.

As subjects entered the experiment each member of the pair was given one from a pair of cards at random. One card showed a 'wave' design and one showed a 'triangle'. These cards were used to indicate roles in the experiment that followed. For the first and second sections of each experiment, individuals worked separately from their partners. In the first section each individual faced n tasks, each comprising a choice between a pair of lotteries; in the second section each individual was asked to make predictions about the n choices their partner had made in section 1. In the third section partners worked together and made joint choices between n pairs of lotteries.

Figure 1 illustrates a typical question. Each depiction of a lottery was composed of three elements: ranges of numbers along the top; payoffs for the subject below and then payoffs for their partner. The numbers along the top corresponded to numbers in the bag of one hundred discs that were used to execute choices at the end of the experiment.

Question 2

Option A

For numbers:	1-50	51-100
You receive	**£20**	**£0**
Your partner receives	**£0**	**£20**

Option B

For numbers:	1-50	51-100
You receive	**£20**	**£40**
Your partner receives	**£0**	**£20**

I choose (***tick one***): Option A ☐ Option B ☐

Figure 1. A Typical Question.

To determine payoffs a question number from 1 to 3n was selected at random in section 3 of the experiment. Outcomes were then as follows:

1. Numbers from 1 to n: Triangle received his or her choice for that question. Wave partner received fifty pence for each correct prediction she or he had made.[5]
2. Numbers from n+1 to 2n: Wave partner played his or her choice for that question. Triangle partner received £0.50 for each correct prediction she or he had made.
3. Numbers from 2n+1 to 3n: the couple played their choice for that question; no money paid for predictions.

To play a choice, the relevant person would choose a chip from the bag of 100 numbered chips and this would yield the corresponding payoffs for them and their partner.

This scheme is incentive compatible if individuals are selfish in their choices, but at least theoretically, an altruistic individual might view the first two sections of the experiment as an exercise in co-ordination in which they should make their choices so as to maximize the predictive success of their partner. To minimize this kind of belief we did not reveal the detailed nature of the experiment until after subjects were separated from their partners. Secondly the prediction section always came after the choice section and subjects were not informed about the type of tasks in section 2 until they had all completed section 1. Finally, the payments for prediction were low. As a benchmark example, a risk neutral income pooling agent would never gain from switching choices in order to improve the predictive success of their partner.

To test the impact of transparency on behaviour we had two information conditions, *private* and *public*. In the private condition, used in experiments A and B, subjects were separated from one another at the start of the experiment and completed sections 1 and 2 in separate rooms.

[5] So wave partners saw questions in the order n+1 to 2n (choice) and then 1 to n (prediction).

Subjects were also not informed of their partner's answers in section 1 of the experiment; nor were they told about the accuracy of their partner's predictions in section 2.[6] Moreover they were informed of these facts. If the question number chosen at random was from section 3, then the experimenters oversaw the choosing of a disc from the bag and payouts in situ. However, if the number was from section 1 or 2 then, to preserve confidentiality, partners were dealt with separately, beginning with the chooser. He or she would go to a table at the far end of the room where they would be reminded of their choice and a disc would be drawn from the bag. Following this he or she would receive any payments owing to them in an envelope and then asked to leave the room, while their partner was dealt with. Again any money owing to the partner was placed in an envelope to protect confidentiality.

For the public condition we removed these barriers: subjects sat separately for the individual questions sections, but they sat in the same room. They were told that their answers were available to their partners and all payments were made in front of both members of the couple. Thus there was full observability of choices between partners.

At the end of the instructions for the first section we placed two questions designed to test understanding. Once all subjects had answered these questions satisfactorily, which nearly all did at the first attempt, they were instructed to answer the choice questions in their own time.

Subjects were recruited from various rural and suburban locations in Surrey and Norfolk, UK, via email, word of mouth, through community groups and using posters. Session sizes varied from two to ten couples and were held at a variety of venues, including a rural village hall and the experimental economics laboratories at the University of East Anglia and Royal Holloway, University of London. All sessions were pen and paper based and we used a script to deliver the

[6] No individual achieved perfection in their predictions or scored zero.

instructions. We required all individuals to be over 21, to be living with their partners and to have been together as a couple for at least one year. We asked subjects to bring evidence of their relationship such as bills with both names on them or photographs and we made random checks.

IV Results.

In all 116 couples took part in experiment – 42 in A, 34 in B and 40 in C. Average payoffs were just under £18 (approximately US$36) per individual, which is more than twice the median hourly post-tax wage for a UK adult. For A and B, standard statistical tests indicated no significant differences between data from questions common to both sub-samples. Consequently, the data from A and B is pooled in what follows.

Table 2 provides background information, based on answers to some socio-economic questions which subjects faced at the end of section 2. The data is presented on an individual basis (N=232) because that was the way it was collected. As can be seen, the range of ages and years together was quite diverse. All the couples in our sample were heterosexual with seventy-three percent married to their current partner. The distribution of children per couple was predictably bimodal with peaks at zero and two. For the financial question, the most common 'other' arrangements were of two kinds: one involved three accounts - one for regular household bills and separate accounts for personal expenditure. The other common arrangement was a decentralised system for routine spending, with each partner responsible for some household bills. This was often coupled with joint decision-making for major or idiosyncratic events.

Table 2. Sample Characteristics.

	Mean	Standard deviation	Range
Children	1.1	1.19	0-4
Age (years)	37.3	10.74	22-70
Years living together	11.1	10.59	1-46
Married	0.71	-	-

Main results.

We turn to the central results. In what follows, the tasks are labelled - their parameter values can be found in Appendix A. In each case, the *number* refers to the task. The numbers attached to tasks do not correspond to the order in which any subjects saw them, but are there simply for reference purposes. For the *letters*, *T* indicates a task faced by triangle subjects, *W* stands for tasks faced by wave subjects and *J* indicates tasks put jointly to couples in the final part of the experiment. *Groups* are collections of tasks which are identical from the viewpoint of an income pooler. A *set* is a triplet of tasks which are identical in outcomes, but differ in the identity of the decision-maker.

As a preliminary to what follows, we first note that subjects rarely choose stochastically dominated options. Specifically, when triangle subjects faced the choice (T18) depicted in figure 1, 6 out of 116 chose the dominated option. The corresponding figures for couples (J18) was 2 out of 116 while for Wave, (W18), 9 out of 116 chose the dominated option for the same task. In W20, a different task, 2 out of 42 wave subjects chose the dominated option. Overall therefore, the stochastically dominated option was chosen on 5% of occasions.

Hypothesis 1. **Information makes no difference to behaviour.**

We compare behaviour across the private and public variants. Figure 2 summarises the results for 10 tasks faced individually. The vertical scale is the frequency of a given option which as can be seen varies little between the variants.

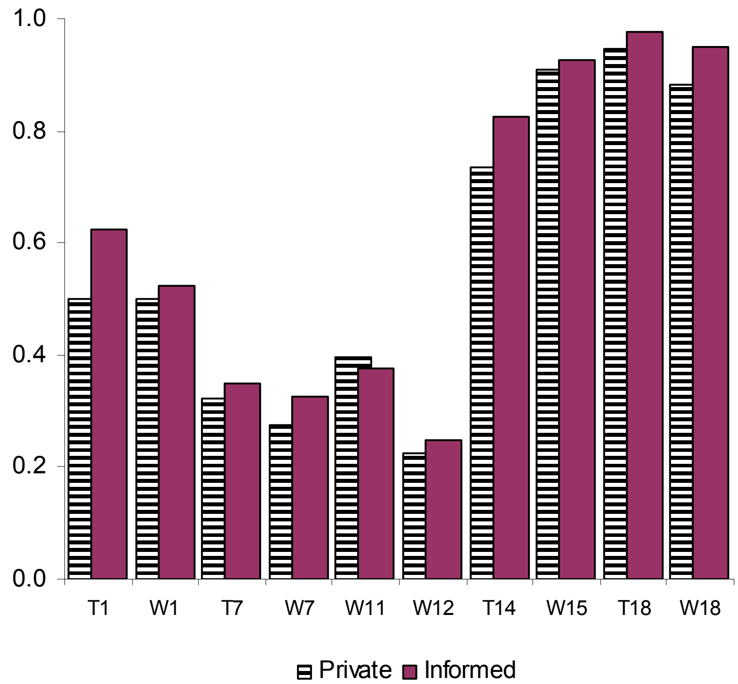

Figure 2. Information and individual choice.

Taking each of the pairs separately, we test whether information makes no difference. The lowest probability with which the null is accepted is 0.138 (Fisher's exact test) for T1. For all but the last two tasks and the first task in the chart there is a one sided alternative hypothesis. The difference between conditions matches the direction of the alternative hypothesis for 5 tasks and is in the opposite direction for 2 tasks. If we simply pool the data from these 7 tasks, the difference between conditions is still not significant (p=0.167, Fisher's exact test). Thus is there no evidence that privacy makes a difference.[7]

Hypothesis 2. Choices are unanimous.

Recall that, by unanimous we mean that individual choices agree and also agree with the corresponding jointly-made choice. From 602 cases from complete sets where we have data of this type, unanimity is present in 281 (46.5 per cent), including the stochastic dominance questions. Seven complete sets of tasks are faced by 74 couples, but none are unanimous in all of them - six are unanimous in six tasks. For the 42 couples who face just two complete sets, 5 are unanimous in both.

To do formal tests, we need to make some assumptions about the stochastic process underlying choices. If agents are assumed to choose without error then any deviation from that predicted by the theory would imply rejection of the maintained hypothesis. This seems too extreme, so in keeping with other models of choice in an experimental setting, we posit some degree of agent error. Specifically, we suppose that the choice of stochastically dominated options represents only the results of a 'tremble' in which subjects mean to choose the dominant option but accidentally pick the dominated lottery. On this basis, a benchmark figure is the 5 percent average for the choice of the dominated option in the four tasks above.[8] Now, where choice is more evenly poised we might expect a greater degree of randomness than that estimated from tasks where one option dominates. An error rate of x implies a rate of inconsistency of 2x(1-x) in

---

[7] As an aside we also do not find differences in the behaviour of the joint choosers. Differences would be unexpected given that the format of section three is unchanged between treatments, but it might conceivably occur if information changed behaviour in the earlier sections of the experiment and this anchored later responses.

[8] For simplicity and robustness, we use the same 5% in all cases, but the figure for couples is borderline significantly different from that for individuals and, arguably 'two heads are better than one' when it comes to spotting errors.

repeated pairwise choice, so that an error rate of 5 percent means an expected inconsistency rate of approximately 9.5 per cent. Looking at the experimental literature, Harless and Camerer, 1994, suggest a somewhat higher range of 15-25 percent for inconsistencies in repeated choice, which implies error rates of 8 to 14.6 percent. In what follows we use the 5 per cent rate, but we also use the Harless and Camerer range as a sensitivity check.

When we test for unanimity, even if we use the 14.6 per cent error rate then we would expect 376 instances of consistency or 62.4%. With this error rate, the null hypothesis of unanimity is still rejected (binomial test, p = 0.000).

Hypothesis 3. Joint choice respects IP.

We have two types of tests: within task and between task. For within task tests, one option strictly stochastically dominates the other from the viewpoint of an income pooler. For between task tests, the choices are equivalent from the income pooler's perspective and so should lead to the same choices. Table 3 shows first the results of within task tests. The IP option is almost always chosen. In fact, the data is close to that which would be predicted with a tremble of 5%.

Table 3. Tests of Income Pooling with Joint Choice

	Within task			Between task				
Task or tasks	J14	J22	J23	J1, J2	J4, J5	J7, J8	J11, J26	J27,J28
N	76	42	42	34	34	76	40	40
Choose option A, first task, %	93.4	92.9	95.2	29.4	88.2	67.1	42.5	77.5
Choose option A, second task,%	-	-	-	29.4	94.1	69.7	45.0	47.5
McNemar's test (d.f. = 1)[9]	-	-	-	0.0	1.0	0.29	0.033	10.3***

[9] McNemar's non-parametric within subjects test does not rely on any particular theory about error rates, except that error rates are constant between tasks.

The remaining columns show five between-task tests drawn from five groups of questions. We test the hypothesis that the proportion jointly choosing the IPE option is the same for each task. With the clear exception of the final pair of tasks, the null hypothesis is not rejected. The *lowest* p value is 0.38. With the final pair, income pooling is strongly rejected.

Hypothesis 4. Individual choice respects IP.

In the case of individuals, the majority of choices are also compatible with IP, but there is also a substantial number who do not choose according to the theory. Table 4 summarises the results of within task binomial tests of the null that the proportion choosing the IP dominating option is 95%. With the exception of W20 we reject the null and as a result we reject the notion that individuals always choose according to IP.

Table 4. Within Tasks Tests of Income Pooling – Separate Choice

Task.	T14	T16	W15	W20	T30	W30
N	116	42	116	42	40	40
Option predicted by IP chosen (%)	79.3	88.1	89.7	95.2	77.5	87.5
Absolute value of z-test	7.75	2.05	2.64	0.071	5.08	2.18
Significance	***	**	***	-	***	**

, * indicates statistical significant at 5% and 1% level respectively.

To shed more light on IP within individual choice, we give an overview of the between task data. The upper part of Table 5 presents ten tasks where for one agent in the couple the payoffs in the question were the mirror image of the payoffs faced by the other partner.[10] Because both lotteries in the task were mirrored, in most cases there is no clear alternative hypothesis to the null that the percentages chosen by each person were the same. In the top half of the table, the between subjects data suggests no significant deviation from income pooling. The results of the within subjects tests are more strongly against the null, though still not overwhelming, with the exception of the pair W11/W12 pair of tasks. It is noteworthy that this is the one pair where the alternative hypothesis is one sided: for this pair, the payoffs to each person in lottery 1 are identical, so mirroring has no consequence. But for the other part of the task, the payoffs going to the chooser in one version first-order stochastically dominates the payoffs going to the chooser in the other version . The switch in the pattern of choices for this pair of tasks is then compatible with choosers placing greater relative weight on their own income.

[10] It is worth emphasising that in the top half of this table we are not proposing that each partner should make the equivalent choice under the null. Rather we are using the fact that individuals were labelled Wave or Triangle randomly and then testing whether the proportion of the sub-samples who prefer lottery 1 to lottery 2 is the same.

Table 5. Between Task Tests of IP in Individual Choice – Mirrored Tasks.

Between subjects										
Task	1	2	4	7	9	11	12	13	32	35
Triangle %	48	53	76	63	50	47	42.5	59	42.5	52.5
Wave %	51	71	76	72	53	35	25	47	45	52.5
N	116	34	34	76	34	34	40	34	40	40
McNemar's test	0.191	3.56*	0	1.49	0.14	1.14	2.88	0.88	0.04	0
Within subjects:										
Tasks and %	W1	51	T7	63	T9	46	W11	39		
	W3	55	T8	71	T10	57	W12	23		
N		76		76		76		116		
McNemar's test		1.33		2.25		3.56*		10.1***		
IP Consistent (%)		84		79		76		72		

% is the percentage choosing lottery 1. For the between subjects data, Tasks should be read as T1, W1 etc. *, **, *** indicates significant at 10% level, 5% and 1% level respectively.

The lack of a departure from IP in the majority of between task tests presented in Table 5 may therefore be due to heterogeneity in preferences and in responses to the differences between tasks. Figure 3 presents information from three groups of IP equivalent tasks where the alternative hypotheses have clear predictions. In this figure we have data from 3 groups of tasks: 3, 5 and 9. Each task is a labelled node. The fraction at each node is the number choosing lottery 1 from the relevant sub-sample. Solid arrows from one task to another point to the task where for the chooser, lottery 1 first order stochastically dominates and/or lottery 2 is first order stochastically dominated. Broken arrows represent transformations that lower lottery 1's risk for the chooser and/or raise lottery 2's risk. As can be seen, in Group 9 nothing much changes as the tasks are transformed: around 50% of the subjects choose lottery 1 whatever the format. For Group 2 there is a clear switch towards the option that is more favourable to the chooser. McNemar's test statistic for this group is 4.47, which is significant at the 5% level. In Group 5, all the changes in proportions between tasks adjacent in the diagram are in the direction predicted by the alternative hypotheses, though only the difference between W12 and W32 is significant (McNemar's test, p-value= 0.011).[11] Summing up, there is some deviation from IP in a manner consistent with individuals placing more weight on their own payoffs than on their partner's income.

Using all the individual data, we examine hypotheses econometrically. The following random utility model is estimated using a probit model with errors clustered at the level of the individual:

$$U = \beta_0 + \beta_{11} M1 + \beta_{12} M2 + \beta_{21} Stdevself + \beta_{22} Stdevpartner + \beta_3 Rho$$

According to this model, the utility of an option is evaluated according to the mean of its payoff for the choosing individual (M1) and his or her partner (M2), the standard deviation for the chooser (Stdevself) and partner (Stdevpartner) and the correlation between the payoffs for Wave and Triangle (Rho). Prior beliefs are that the coefficients on the mean variables are positive and those on standard deviation and correlation are negative. To estimate the model we take differences for each variable between option 1 and 2.

Table 6 summarises our results for four slightly different specifications. In this table, the dependent variable is choosing option 1. The reported coefficients in the table are marginal values with z-values in parentheses. Right hand side variables ending in 'self' describe features of the payoffs for the choosing individual; variables ending in 'partner' describe features of the payoffs for the non-choosing partner. For all specifications reported we include interaction terms between M1, M2 and whether the subject is male. For some specifications we also include interaction terms for mean payoffs and the number of children and years spent together as couple. We also

[11] As can be seen from Table 4, the W11-W12 difference is also significant on the whole sample.

try experimenting with other recorded variables such as age and household financial arrangements, but these produce no sign of statistical significance and are not reported.

Table 6 Econometric testing for income pooling.

	(1)	(2)	(3)	(4)
Mean for self (M1)	0.022	0.0182	0.022	0.021
	(10.12)***	(9.72)***	(10.12)***	(8.48)***
Mean for partner (M2)	0.021	0.0191	0.021	0.020
	(9.20)***	(9.02)***	(9.32)***	(7.85)***
Male x M1	0.0069	0.0069	0.0069	0.010
	(3.20)***	(3.22)***	(3.20)***	(3.16)***
Male x M2	0.0032	0.0032	0.0032	0.0056
	(1.40)	(1.41)	(1.40)	(1.66)
Years*M1	-0.00031		-0.00036	-0.00036
	(3.60)***		(3.61)***	(3.61)***
Years*M2	-0.00029		-0.00032	-0.00032
	(3.05)***		(3.05)***	(3.02)***
Children*M1	-0.00082			
	(0.76)			
Children*M2	-0.00055			
	(0.45)			
Rho	-0.044	-0.049	-0.045	-0.016
	(2.06)**	(2.28)**	(2.07)**	(0.53)
Stdev. Self	-0.0046	-0.0046	-0.0046	-0.0033
	(4.04)***	(4.07)***	(4.04)***	(2.16)**
Stdev. Partner	-0.0037	-0.0037	-0.0037	-0.0051
	(3.23)***	(3.21)***	(3.23)***	(3.23)***
Male x Stdev. Self				-0.0026
				(1.18)
Male x Stdev. Partner				-0.0028
				(1.26)
Male x IP Safe				-0.059
				(1.37)
Observations	2616	2616	2616	2616
PLL	-1647	-1654	-1647	-1645
Test 1 “all”	0.000	0.000	0.000	0.000
Test 2 “Female Means”	0.619	0.890	0.660	0.733
Test 3 “Male means”	0.045	0.038	0.043	0.042
Test 4 “Std. dev.”	0.596	0.572	0.595	0.460

1. Values are marginal effects. Figures in parentheses are t-statistics based on clustered standard errors. *, **, *** = significant at 10% level, 5% level, 1% level respectively.
2. PLL = log pseudo-likelihood
3. Figures for tests are p-values.

Equation (1) includes interaction terms for children and years together. Equation (2) drops both of them and equation (3) has just the interaction terms for years. Equation (4) also allows gender to affect attitudes to risk and correlation. As can be seen there is not much evidence that attitudes to risk are affected by gender and there is also little evidence that the number of children affects the weight put on mean payoffs.

At the base of the table are the results of four tests. The first is a likelihood ratio test with the null that the model as a whole has no explanatory power. The second test is for the equality of the coefficients on the two mean difference variables. Since there is also an interactive dummy for men, Test 1 is a two-sided Wald test of the hypothesis that female subjects weight their own payoffs and their partner's payoffs equally. This null is always accepted. The third test uses the delta method to test whether the relative weight that women put on their own payoffs (versus their partner's) is equal to the relative weight placed by men. In all the reported specifications this null is rejected at the 5% level. To put this in perspective if we take equation (1) for a woman with zero years together and no children, the ratio of the M2/M1 coefficients in equation 3 is about 0.97. For men the corresponding ratio is 0.84. As the years together increases the gap between the genders gets bigger. At 20 years, the ratio is 1.00 for women and 0.82 for men and when years together is 40, then the relevant ratios are 1.08 and 0.79. When two children are present these ratios change only fractionally. Thus in our sample, men discount their partner's payoffs by 15-20%. Women though put approximately equal weight on their partner's payoffs.[12] The final test in the table is for equal weight on standard deviations and here there appears to be no evidence that the subjects systematically weight their partner's risk differentially to their own. Moreover, as equation (4) shows, there is no obvious gender difference in the relative weight attached to risk.

---

[12] These econometric tests are subsidiary to the direct tests of IP discussed above. There we find no difference between the genders: females are just as likely as males to choose IP dominated outcomes. Thus we do not claim that women always pool income.

As a check we also estimate the model for joint choices, where 'self' becomes triangle payoffs and 'partner' becomes Wave payoffs. As one would anticipated we find no difference in coefficients according to identity. When, instead of triangle and partner we use male and female as labels, again we find no difference in coefficients.

Hypothesis 6. **Choices respect interpersonal separability.**

Table 7 shows the results for all of the tasks for which Interpersonal Separability makes a clear prediction of one option, but where for the household as whole one alternative does not stochastically dominate the other. Viewing each individual separately, each task shares the feature that one option first order stochastically dominates the other lottery. In four such tasks, for one partner, options A and B are identical while for the other partner one option strictly dominates the other. For W5 and J5, one option strictly dominates for both individuals. As can be seen, in none of the tasks does the proportion which chooses the option predicted by the theory rise above 50%. This is true whether the choice is made by individuals or by couples.

Table 7. Within task test of Interpersonal Separability properties.

Task	J19	T2	W2	J2	W5	J5
N	76	34	34	34	34	34
Instances predicted option chosen (%)	31.6	47.1	29.4	29.4	29.4	11.8
Chi-squared.	493***	124***	236***	236***	236***	383***

*** statistically significant at 1 per cent level, d.f. = 1.

The chi-squared values in Table 7 are calculated using the 5 percent error rate figure. Clearly the null hypothesis is rejected for each task at any widely recognised level of statistical significance and the conclusion is not sensitive to the choice of error rate. All of the tasks in Table 7 share one key feature: the alternative, apparently dominated option is risk-free when viewed by agents who satisfy IP. In the vast majority of cases, it is this 'safer' option which is chosen by the subjects. This accords with the econometric results in Table 6, where the coefficient on the correlation between partners' payoffs was negative and statistically significant.

***Hypothesis** 7*. **Joint choices are Pareto efficient.**

According to P, if both individuals strictly prefer lottery A over B or one individual is indifferent and the other strictly prefers A, then as a couple they should not choose B. Figure 3 shows the pattern of agreement within couples for eleven complete sets of identical tasks in the experiments. For each complete set, the first bar shows the number of couples where individual choices differ. The second bar shows the number of cases where individual choices are congruent and are the same as the choice made jointly. Leaving aside for the moment the possibilities of indifference or error, if P holds there should be no cases where the joint choice differs from the individual choices, when those individual choices are the same. However, as the last series of bars shows, in each set there are couples where joint choice does have this property. On average, 23 per cent of couples with identical individual choices plump for a different option in joint choice.

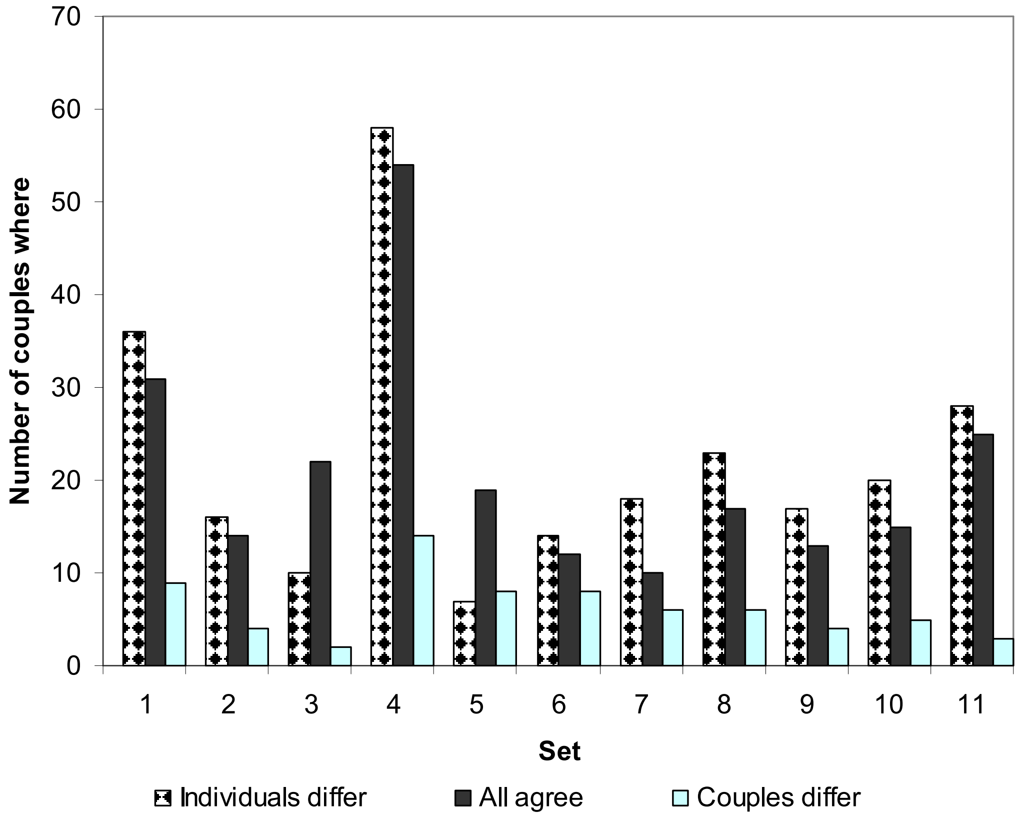

Figure 3. Agreement of preferences.

One theoretical possibility is that in all the relevant cases both individuals are indifferent between options A and B, so that the joint choice does not represent a rejection of the Pareto principle. An alternative and perhaps more realistic theory is that of other regarding behaviour combined with an error in prediction which is then corrected once individuals are able to communicate with their partners. However there is no supporting evidence for this hypothesis in the prediction data. Another possibility is that the pattern could represent the result of the random element in preferences. We conduct chi-squared tests of this possibility, based on the 5 per cent benchmark. For all but complete set 3, the null is decisively rejected at a 1 percent level.

There are reasons to show some scepticism about the test in this context. First, if all choices are subject to a tremble of 5 percent, then some of the apparent agreements between individual choices used as the basis of the test will actually reflect previous errors. Dependent on how disagreement on preferences is resolved this may raise or lower the expected rate at which joint choice should diverge from individual choices. Secondly, recall that the 5 percent figure is based on data from tasks where one option stochastically dominates the other and therefore might underestimate the rate of error in more evenly poised decisions. If we use the lower of the error rates (8 percent) from Harless and Camerer's 1994 range then the conclusions are unchanged, but if we use the upper value (14.6 percent) then the null hypothesis would be rejected for only the last three data sets at 1 percent levels of significance.

However, there is a clear pattern of change between individual and joint choice: the reversals of preference are not random. For complete set 4, where neither option is definitively safer than the other, six couples switch in one direction and eight in the other. For the remaining sets, where in each case one option can be ranked as the safer option, 40 of the switches are towards the safer option in joint choice and only fifteen are towards the risky option. In fact, it is set 10 where more couples switch to the riskier option (3 versus 2 switching to the safe option). So, apparently couples are more risk averse in their behaviour than individuals. In some cases this manifests itself in a complete switch from the riskier option chosen by both individuals to the safer option. This is possibly the most surprising result of the experiment.

The wider evidence for the claim that couples are more risk averse than their component individuals is summarised in Figure 4 which shows the proportion choosing equivalent options for ten complete sets of tasks where there is data on choice from both individuals and also from their joint choice as a couple and where it is possible to rank the riskiness of the options from the viewpoint of an agent satisfying IP. For complete sets 1-3, 6-11 one option has no risk for the income pooler. In all but set 2 one option is completely risk free for both individuals. In set 2 it is not possible to order unequivocally the riskiness of the options from an individual perspective.

For complete set 5, both options involve risk, but one option is safer.

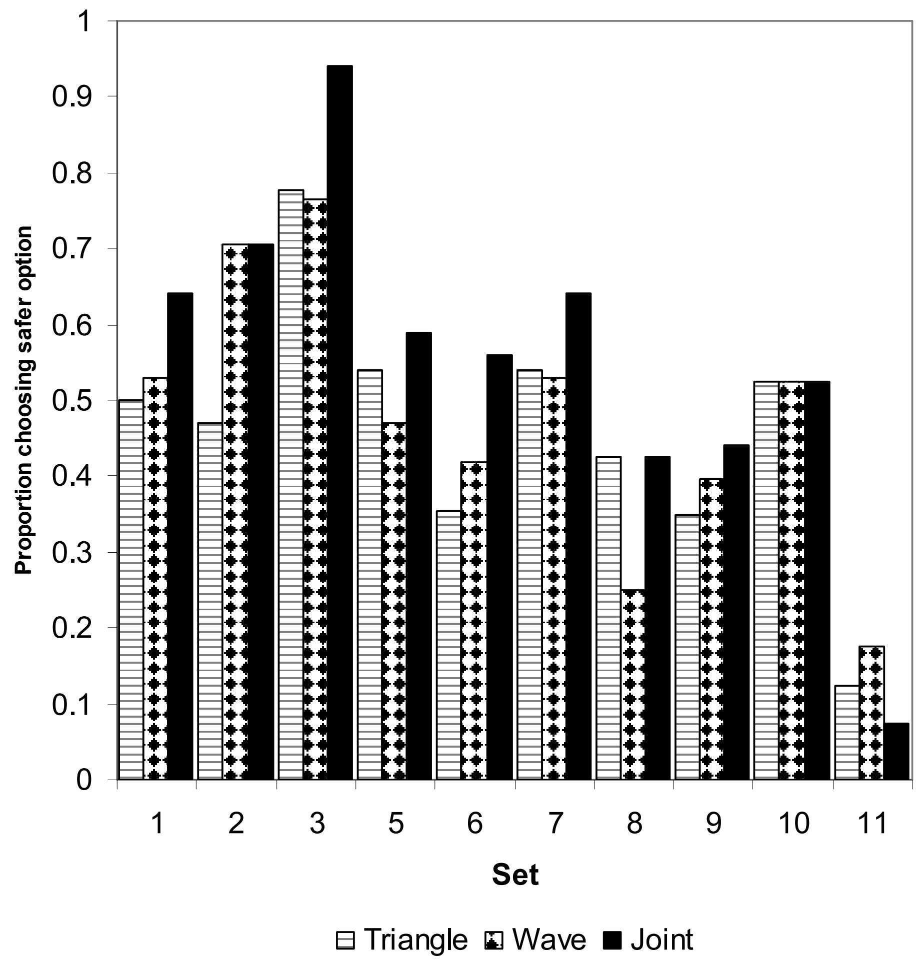

Figure 4. Proportions choosing equivalent option.

For all sets, the bars show the proportions choosing the IP safer option. With the exception of set 11, the bar is highest for joint choice. Using a signs test, the triangle proportions are significantly lower than the jointly-chosen proportions for sets 1 (p=0.020), 2 (p=0.037), 3 (p = 0.01), and 6 (p=0.023). On the equivalent test for wave versus jointly-chosen proportions the significantly different sets are 1 (p=0.020), 3 (p=0.005), 5 (p=0.044), 7 (p=0.040) and 8 (p=0.033).

It is not clear why couples should tend to be more risk averse than individuals. As the survey by Kerr et al, 1996, makes clear, the psychological evidence is that the relationship between individual and collective decisions is not a simple one. Collective decisions may be more or less prone to errors of judgement depending on the nature of the task, but one consistent pattern is that collective decisions are often more extreme than individual choices. However, here it is not obvious that the safe option should be viewed as more extreme. Hypotheses 4 and 5 suggest that couples behave more like income poolers than individuals. Possibly therefore, the safety of one of the options is simply more salient when decisions are made jointly. Alternatively, perhaps safety is more attractive in joint decisions because choosing the safer option minimises the risk of recrimination once the state of the world is revealed. One possibility we can discount is that individuals routinely underestimate their partner's degree of risk aversion, which is then corrected when they are asked to make joint choices. The relevant prediction data for the sets in figure 5 shows no systematic tendency to over- or under- predict the degree of risk aversion. One other possibility lies in the negotiation opportunities opened up by joint decision-making. When making individual choices, subjects have to make conjectures about ex-post trades, which might be false; they are also denied the possibility of ex ante trade. However, given the opportunities for risk-sharing provided by negotiation, this kind of explanation would sit more easily with individuals making more risk averse choices compared to those made by couples. We are left with more speculative hypotheses – perhaps when placed together couples are more conscious of being blamed for choosing the risky option if it fails. If 'blame aversion' is responsible for the difference then we might expect it to disappear in the public variant, where subjects know that their choices will be available to their partners. Note though that we get this pattern of difference between individual and joint choice in both the private and public experiments.

V Discussion.

This paper reports on an experiment designed to test household models. While there are many questions that remain to be answered we have shown that experiments can be used to test theories of the household in the same way that they have been used to test theories of individual decision-making. In our experiment, when couples make joint choices, income pooling is a typical feature of their decisions, but when partners make decisions separately they place less weight on their partner's payoffs. Econometrically, the discount is estimated as £0.16 for men and very close to zero for women. Direct tests of IP however, show that both sexes show some tendency to choose the IP dominated option when it is individually disadvantageous. Nearly always individuals and couples pay some attention to the correlation of payoffs, so that separability is not a common feature of choices. The choices made by individuals often differ from the choices made by couples, so that unanimity is also rare. Although many choices do respect the Pareto principle, a significant fraction (about 23 per cent of the relevant cases) do not, particularly when one option is clearly riskier than the other. Unlike Ashraf's study of Filipino households in our UK sample, whether choices are observed by the other partner makes little difference to behaviour.

Taken at face value, the results suggest that no standard economic model of the family has universal support. Two models in particular do not fare well. A simple weighted-average-of-utility model with no ex post transfers performs particularly badly. Secondly, the strictest version of the unitary model - in which we see all choices, joint and individual, coincide – is not consistent with the data for more than a few households.

References.

Alderman Harold, Chiappori P.A., Haddad L, Hoddinott J, Kanbur R, 1995, Unitary Versus Collective Models Of The Household - Is It Time To Shift The Burden Of Proof, *World Bank Research Observer* 10 (1): 1-19 Feb.

Ashraf, N. 2005, Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines, Harvard University.

Bateman, I. and A. Munro, 2005, An experiment on risky choice amongst households, *Economic Journal*, 115, C176-C189.

Bolin, K., L. Jacobson and B. Lindgren, 2002, The family as the health producer - when spouses act strategically, *Journal of Health Economics*, 21: 475-495.

Becker, Gary S., 1974, A theory of social interactions, *Journal of Political Economy*, vol. 82 (6), 1063-1093.

Browning, Martin, and Pierre-Andre Chiappori, 1998. Efficient intra-household allocations: a general characterisation and empirical tests. *Econometrica* 66 (6), 1241–1278.

Chen, Zhiqi and Frances Woolley, 2001, A Cournot-Nash Model of Family Decision-making, *Economic Journal,*111, 722-748.

Corfman, Kim. P. and Lehmann, Donald. R., 1987, Models of cooperative group decision-making and relative influence: An experimental investigation of family purchase decisions. *Journal of Consumer Research*, 14: 1-13.

Doss, Cheryl, 1996, Testing among models of intrahousehold resource allocation, *World Development*, 24(10), 1597-1609.

Harless, David W. and Colin F. Camerer**,** 1994, The Predictive Utility of Generalized Expected Utility Theories, *Econometrica*, Vol. 62, No. 6. 1251-1289.

List, J. and Steven D. Levitt , 2005, What Do Laboratory Experiments Tell Us About the

Real World? University of Chicago September

Lundberg, Shelly J. and Robert A. Pollak, 1993, Separate Spheres Bargaining and the Marriage Market, *Journal of Political Economy*, Vol. 101, No. 6. Dec., 988-1010.

Lundberg, Shelly J., Robert A. Pollak and Terence J. Wales, 1997, Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit, *Journal of Human Resources*, Vol. 32, No. 3. (Summer,), pp. 463-480.

Kerr, Norbert L., MacCoun, Robert J.; Kramer, Geoffrey P. 1996 Bias in judgment: Comparing individuals and groups. *Psychological Review*. Oct Vol 103(4) 687-719

Pahl, Jan, 1990, Household spending, personal spending and the control of money in marriage, *Sociology*, 24:1, 119-138.

Peters, H. Elizabeth, A. Sinan Ünür, Jeremy Clark, William D. Schulze, 2004, Free-Riding and the Provision of Public Goods in the Family: An Experimental Test of the Rotten Kid Theorem, *International Economic Review*, 45(1), 283-299

Phipps, Shelley A., and Peter S. Burton., 1998, What's Mine is yours? The influence of male and female incomes on patterns of household expenditure, *Economica*, 65, 599-613.

Rothschild, Michael and Joseph E. Stiglitz, 1970, Increasing risk I: A definition, *Journal of Economic Theory* 2: 225-243.

Udry, Christopher, 1996, Gender, Agricultural Productivity and the Theory of the Household, *Journal of Political Economy*, 104/5.

Udry, Christopher, and Esther Duflo, 2001, Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices, Yale University *mimeo*.

Warr, Peter G., 1983, The private provision of a public good is independent of the distribution of income, *Economic Letters,* 13, 207-11.

Appendix A. The tasks.

Group	Experiment	Lottery 1				Lottery 2			
		Triangle		Wave		Triangle		Wave	
		£20	£40	£20	£40	£20	£40	£20	£40
1	T1, W1, J1, (Set 1)	1-100	-	-	-	21-70	71-100	-	-
	T2, W2, J2, (Set 2)	21-100	-	1-20	-	21-100	-	71-100	-
	W3	-	-	1-100	-	-	-	21-70	71-100
2	T4, W4, J4, (Set 3)	1-100	-	-	-	-	-	-	41-100
	W5, J5	1-50	-	51-100	-	41-100	-	41-100	-
	T6	1-100	-	-	-	-	41-100	-	-
3	T7, W7, J7, (Set 4)	31-100	-	1-100	-	21-70	71-100	-	21-100
	T8, J8	1-100	-	31-100	-	-	21-100	21-70	71-100
4	T9, W9, J9, (Set 5)	-	-	51-100	-	-	-	-	71-100
	T10	51-100	-	-	-	-	71-100	-	-
5	T11, W11, J11, (Set 6)	1-100	-	1-100	-	1-100	-	21-70	71-100
	T12, W12, J12, (Set 8)	1-100		1-100	-	21-70	71-100	1-100	-
	T32, W32, J32, (Set 9)	1-100		1-100		1-20	21-70 then £60		
	W34,		1-100			1-20	21-70 then £60		
6	T13, W13, J13, (Set 7)	-	-	1-100	-	-	1-70	-	-
7	T14, J14	51-100	-	1-50		1-100	-	-	71-100
	W15	31-100	-	1-30	-	-	71-100	1-100	-
8	T34		1-100			1-60		61-100	1-60
	T35, W35, J35, (Set 10)	1-100		1-100		1-60		61-100	1-60
	T36, J36	1-100		1-100		1-40	Then £60		
9	T9, W23, J28, (Set 11)	1-100		1-30		1-100			1-20
Other	T16	51-100	-	31-100	-	1-100	-	-	71-100
10	T18, W18, J18 (Set 12)	1-50	-	51-100	-	1-50	51-100	51-100	-
	W20	1-70	-	21-70	71-100	71-100	-	1-70	-
	J22	-	71-100	1-70	-	1-60	61-100	-	-
	J23	-	-	21-70	71-100	1-40	-	-	71-100
	T30, W30	1-100		1-100			1-80 then £60		

Note: numbers show the ranges of disc values for which the corresponding payoffs were awarded. In a few tasks, a payoff of £60 was available. 'then £60' means higher lottery numbers yielded £60. We omit numbers for states of the world where the payoff was 0. Groups 1-10 represent collections of tasks which are IP equivalent. Sets 1-12 are collections of tasks which are identical.

Appendix B. Example instructions. Not for publication.

Subject pair___Partner_Δ_

Section 1.

- In this section you will face a series of twelve different choices involving options like example 1 shown below.
- Each choice will involve two options, labelled A and B.
- For each question your task is to choose the option you would prefer to have.
- At the end of the experiment, one question number from 1-36 will be drawn at random.
- If one of these questions is selected at the end of the experiment you will play out the option chosen by you.
- You will be asked to draw a number from 1-100 from a bag containing 100 discs.
- You will receive your prize corresponding to that number and your partner will receive their prize corresponding to the same number.
- Your choices are confidential - we will not reveal them to your partner. (In experiment C: **Your answers will be known to your partner**)

The options you will face are similar to the one shown in Example 1.

Example 1.

For numbers:	1-20	21-70	71-100
You receive	**£0**	**£0**	**£40**
Your partner receives	**£0**	**£20**	**£40**

The row beginning ‘For numbers’ has three ranges of numbers, corresponding to the numbers in the bag. Below each range is a prize for you and then a prize for your partner.

For instance, under the range 21-70 you can see a prize of £0 for you and a prize of £20 for your partner. These are the prizes you would get if you had this option and you pulled a number between 21 and 70 from out of the bag.
[please turn over]

Appendix B. Example instructions. Not for publication.

Example 2 shows choices as you will see them. In each case you must choose ONE option by ticking the appropriate box.

Example 2.

Option A

For numbers:	1-60	61-100
You receive	**£20**	**£20**
Your partner receives	**£0**	**£20**

Option B

For numbers:	1-20	21-70	71-100
You receive	**£0**	**£0**	**£40**
Your partner receives	**£0**	**£20**	**£40**

I choose (***tick one***): Option A☐ Option B☐

For instance, suppose you chose Option B by ticking the relevant box, then if the number 25 was pulled from the bag at the end of the experiment, you would receive nothing and your partner would get £20.

Check your understanding.
Suppose that in the example question you had chosen Option B and that this question was picked at the end of the experiment.

What would you receive if the number 80 was picked from the bag? (***tick one***) **£0** ☐ **£20** ☐ **£40** ☐

What would your partner receive if the number 20 was picked from the bag? (***tick one***) **£0** ☐ **£20** ☐ **£40** ☐

To sum up, for each question your task is to choose the option you prefer.