WHY ANOMALIES CLUSTER IN EXPERIMENTAL TESTS OF ONE-SHOT AND/OR FINITELY REPEATED GAMES: SOME EVIDENCE FROM PSYCHOLOGY

by

ANDREAS ORTMANN [a,b] and RALPH HERTWIG [a]

[a]Center for Adaptive Behavior and Cognition
Max-Planck-Institut fuer Bildungsforschung
Berlin, Germany

[b]Center for Economic Research and Graduate Education
Charles University
Prague, Czech Republic

June 13, 2000

Abstract

We conceptualize the decision problem that participants in certain experiments face as a Bayesian updating problem involving the ”typical behavior” of a reference group (participants in the particular experiment) as anticipated baseline behavior and the ”individuating behavior” that can be inferred from the instructions for members of the same reference group. This framing of the decision problems that participants in experiments typically face allows us to draw on evidence from psychology where the base rate fallacy has been studied widely in various domains. Recently, a consensus is emerging among psychologists that the occurrence of the base rate fallacy is highly contextual and that in many well-defined social circumstances base rates are surprisingly potent and reliable even when determined efforts are undertaken to exorcize them. We mine this evidence to shed light on the occurrence of anomalies in contexts such as social dilemma/public good provision/common pool problems, gift-exchange/principal-agent/one-sided prisoners' dilemma games, ultimatum games, trust games, moonlighting games, etc.

Introduction.

Humans typically live in repeated-game contexts and are ”intuitive statisticians” (Cosmides & Tooby 1996): they construct population frequencies of the typical actions (or characteristics) of members of a particular category in a particular context ”through natural sampling” (Aitchison & Dunsmore, 1975; Gigerenzer & Hoffrage, 1995; Cosmides & Tooby 1996). These ”base rates” of actions (or characteristics) will be typically used to predict the behavior (nature) of the members of a particular category in a particular context. In this sense, repeated-game contexts generate base rates that enter decision making under uncertainty.

Most laboratory situations are far removed from everyday experiences. Instructions are instances of ”individuating information” about situations that are atypical in many respects (e.g., a one-shot game or finitely repeated game, abstractness of the laboratory scenario, existence of an experimenter whose agenda may not be obvious, etc.). Most importantly, instructions provide information about the incentives that members of a particular category have in a very particular and, in any case, highly unusual context. This individuating information, if it is understood at all (e.g., Binmore 1999), often leads experimental participants to inferences about behavior that contradicts what base rates suggest; this is clearly the case for all those games where the game-theoretic predictions for one-off or finitely repeated games on the one hand and indefinitely repeated games on the other, differ (e.g., social dilemma/public good provision/common resource problems, gift exchange/principal-agent/one-sided prisoners' dilemma games, ultimatum, trust, and moon-lighting games, etc.).

In the present paper, we review evidence from psychology about the extent and circumstances that allow for inviduating information to overcome base rate beliefs. Contra the proposition -- commonly known as base rate fallacy -- that judgements are dominated by the information extracted from the evidence while prior beliefs are ignored or largely underutilized (Ofir, 1988, p. 343), there is strong evidence that base rate beliefs are firmly entrenched and not easily overcome. In fact, in some quarters the base rate fallacy is now considered ”a myth” (Koehler 1996, p. 5) and a consensus seems to be emerging that everyday experiences typically induce potent and resilient base rates that are not easily unsettled by individuating information and that the base rate fallacy may well-be a myth for most of those situations that people typically encounter in real life (e.g. Koehler 1996; Cosmides and Tooby 1996). Underweighing of individuating information typically happens for every-day base rates (e.g., social or sex stereotypes). Based on this evidence, our working hypothesis is that experimental subjects come into the laboratory with similarly ”potent and reliable” base rates about social contexts, not allowing them to pay due attention to the individuating information about the laboratory situation. We note that many anomalies in economics are connected to experimental designs that try to implement one-off or finitely repeated game or decision scenarios, i.e., situations where experimental participants have to overcome the very potent and resilient base rate experiences that serve them well in everyday life. We then connect these two observations.

We conclude that there is a good chance that many of the experimental results for the classes of games discussed here are artifacts of experimental design. Unfortunately, while anomalies are easy to construct, they are very difficult to de-construct because such deconstruction involves, at least for most of the games considered here, the construction of a counterfactual reality. [On the other hand, the literature on base rates - because it demonstrates both the presence and absence of base rate effects, and seems to do so under well-defined circumstances - may guide us to a better understanding of where anomalies in experimental economics come from and how they can be made to disappear.]

The paper is structured as follows: In section One we discuss the base rate problem and instances of the ”base-rate fallacy”. In section Two we discuss the environments under which they emerge. In section Three we enumerate and classify anomalies that have been reported in the experimental economics literature. In section Four we discuss the environments under which they emerge. In section Five we argue why that many of the so-called anomalies in economics are connected to experimental designs that try to implement one-off or finitely repeated game or decision scenarios (e.g., results from social dilemma games, ultimatum and dictator games, trust games, principal-agent games, etc.), i.e. situations where experimental participants have to overcome the very potent and resilient base rate experiences that serve them well in everyday life. Because it has been shown that it is difficult to deactivate everyday base rates, we conclude that there is a good chance that many of the experimental results for the classes of games discussed here are artifacts of experimental design very similar to the base-rate fallacy.

Section One: the base rate problem and instances of the "base-rate fallacy" (a primer)

> "In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not a Bayesian at all." (Kahneman & Tversky, 1972, p. 450)
>
> "The genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact." (Bar-Hillel, 1980, p. 215)
>
> "Most people do not understand Bayes Theorem, the statistical method of weighing evidence ... Given two sources of evidence, most would not weight them correctly. ... People tend to underweight or altogether ignore fundamental background data. ... " (Salop, 1987, p. 158)
>
> "Much of the research dealing with human judgement under uncertainty in a Bayesian framework has focused on the neglect or underutilization of base-rate information. Judgements are dominated by the information extracted from the evidence while prior beliefs are ignored or largely underutilized." (Ofir, 1988, p. 343)

Bayes' Rule instructs us that in evaluating whether a hypothesis (H) is true relative to its complement (5H) one ought to incorporate both initial beliefs ("priors" from here on) and sample information ("Data", or D, from here on) to get updated beliefs ("posteriors" from here on). Formally,

$$(1)\qquad \frac{p(H|D)}{p(\,5H|D)} = \frac{p(D|H) \times p(H)}{p(D|\,5H) \times p(5H)}$$

where p(H) and p(5H) denote the priors, p(H|D) and p(5H|D) denote the posteriors, and p(D|H)/p(D|5H) denotes the so-called likelihood ratio. The likelihood ratio is the quotient of what is called the hit rate or true positive in the numerator and the false alarm rate or false positive in the denominator. D is sometimes called the individuating, or diagnostic, information; the priors are sometimes called the base rates.[1] Note that (1) is the ratio of two other frequently used versions of Bayes' Rule:

$$(2)\qquad p(H|D) = \frac{p(D|H) \times p(H)}{p(D)} = \frac{p(D|H) \times p(H)}{p(D|H) \times p(H) + p(D|\,5H) \times p(5H)}$$

[1] Unfortunately, although these terms often are used synonymously, the individuating information may not be diagnostic, and priors may not reflect base rates. We'll return to this issue presently.

where (2) is derived from p(D 1 H) = p (D) p(H|D) and p(D 1 H) = p (H) p(D|H) and involves the conditional probability p(H|D) that one would attach to the event H (or, D) if one knew that the event D had already occurred (Binmore 1992, p. 71). Likewise, p(D|H). Likewise, the derivation of p(5H|D). Two examples of an application of Bayes' rule are provided in appendix A.

In applying Bayes' Rule, many things can go wrong. For example, people may pay too little or too much attention to the false alarm p(D|5H) (e.g., Ofir, 1988). Likewise, people may under- or overestimate the hit rate. Most attention has focused on (the ratio of) the priors, i.e. the base rates. For the longest time, there was a wide-spread belief - as exemplified by the four quotations at the beginning of this paragraph - that base rates typically are unduly ignored and individuating information as provided by hit rate and/or false alarms unduly affect people's inferences and decision making. This belief has become known as "base-rate fallacy".

More recently, and culminating in an extensive review of the literature (Koehler 1996; 1996a) plus numerous comments on the target article (Koehler 1996), a consensus is emerging that the base rate fallacy may well be "a myth" and that people in many contexts are rather conservative in their assessment of individuating information, especially if it contradicts their base rate beliefs. Base rate beliefs, in a memorable phrase, can be surprisingly "potent and resilient" (Nelson, Biernat, & Manis 1990) to individuating information. (Interestingly, this brings us a full turn back to the old conservatism idea, e.g., Edwards 1968; Phillips & Edwards 1966. Interestingly, too, how this "emergence of a myth" (Koehler 1996, 1996, p. 5) was possible. For some relevant conjectures, see Lopes 1991, 1992.)

That the base rate fallacy may be a myth does not, of course, establish that conservatism is the default option. Rather, the simple fact is that it all depends (Koehler 1996a, p. 43): There are situations where base rates are underweighted and situations where they are overweighted. Of particular interest here is evidence is that the latter case is particularly prevalent in social judgement. It is one of the merits of Koehler (1996) and its commentators to have amassed evidence that allows us to specify the conditions under which base rates are likely to be overweighted.

Section Two: *... **and the environments under which they emerge.***

> " ... a large body of literature indicates that stereotype base rates are used and overused in social judgement." (Koehler, 1996, p. 43)

Koehler (1996, p. 3-4 and 4-5, respectively) discusses the history of experimental results on the classic and paradigmatic lawyer-engineer problem (Kahneman & Tversky, 1973)[2] as one example motivating his review and studies on social judgement as the other. We focus here on the latter because it is of immediate relevance to the problem of experimental design, implementation, and results discussed below.

Locksley and her colleagues argued initially that sex and other social stereotypes were disregarded when individuating information was made available (Locksley et. al. 1980; 1982). Contra these experimental results, and of particular importance in the current context, Nelson, Biernat, & Manis (1990) established that every day base rates of group stereotypes are, in a memorable phrase, "potent and resilient" (p. 664), having "a continuing, robust effect on the evaluations of individual group members." (p. 671) even though the authors tried hard to exorcize them. (The study was an attempt to de-activate base-rate effects.) In earlier studies, Ginnosar and & Trope (1980, 1987), Krueger & Rothbart (1988), and Hilton & Fein (1989, using a task similar to one in Locksley et al. 1980) demonstrated that base rates are effective when individuating information is inconsistent or irrelevant, or at least not in line with subjects' own stereotypes (Rasinski et al. 1985; also using a task similar to one in Locksley et al. 1980).

This issue of the diagnosticity of the individuating information has (since) received increasing attention. There seems to be consensus that people are sensitive to the diagnosticity of the evidence (e.g. Trope & Bassok 1982; see also Ofir 1988; ...), especially when base rate and hit rate are high and false alarm is either inconsistent with these cues ($p(D|5H) > 50\%$) or redundant ($p(D|5H) < 50\%$).

.

This little review already indicates some of the determinants of the overweighting of base rates. Now in more detail.

---

[2] In Kahneman & Tversky (1973) one set of subjects were (falsely) told that a panel of psychologists had written personality descriptions of 30 engineers and 70 lawyers based on the results of personal interviews and personality tests; the other set of subjects were told that the underlying population was 70 engineers and 30 lawyers. All subjects were then told that five descriptions had been chosen from the respective pools at random. [More here?]

In his review, Koehler (1996) pulls together a ”diverse and as-yet-unsynthesized literature” (lit. cit., p. 5); he then proposes a model of base rate use that focuses on task structure, task representation, and ”normative influences” such as relative diagnosticity of both base rates and individuating information and reliability of the source that communicated them (see his figure 1); Koehler also discusses heuristics that focus attention on the base rate. Our following summary of the key determinants of the overweighting of base rates is informed by Koehler's review. Our discussion is focused on the particular purpose of alerting economists to findings from psychology relevant for the issue at hand. These results are encapsuled in statements (catch phrases) that will be used in our discussion below.

Catchphrase 1: FREQUENTIST PROBLEM PRESENTATIONS

It is by now well-known that human beings in probabilistic reasoning tasks are remarkably sensitive to the format in which information is presented (e.g., Gigerenzer & Hoffrage 1995, Cosmides & Tooby 1996; Brase, Cosmides, & Tooby 1999). For example, Gigerenzer & Hoffrage (1995) found that with frequency representations, subjects arrived at the numerically exact estimate using a Bayesian algorithm (including pictorial equivalents and shortcuts) in about 50% of the cases. And, Sedlmeier & Gigerenzer (1999) report several studies that compare learning success for different scenarios such as teaching people to represent the problems in frequencies and how to insert probabilities in Bayes' Rule. They find that the immediate generalization effect for the representation training was about twice as high as that for rule training. Maybe more importantly, this effect remained stable over follow-up tests (one week, five weeks) whereas performance in the rule-training group showed the typical forgetting curve. Cosmides & Tooby (1996) report even stronger performance effects. These results suggest that humans have inductive reasoning mechanisms that reflect rational reasoning principles if these mechanisms are supplied with the appropriate information format, namely frequency problem presentations. Such presentations will ”automatically” happen if subjects can sample the frequency of certain characteristics/action choices in social situations that are like or similar to those that they encounter in the social situation ”experiment”.

Catchphrase 2: UNAMBIGUOUS SAMPLE SPACES

Koehler (1996, p. 7) stresses that the ”frequentist problem presentation thesis” will find more support when repeated sampling happens from reference classes whose size and composition are stable and known. For the subject pools from which subjects are typically drawn that seems indeed so.[Here footnote on subject pools in economics and psychology - see Ortmann & Hertwig 2000.]

Catchphrase 3: CREDIBILITY OF THE SOURCE

In a typical base rate task in psychology, subjects are provided with a base rate fallacy (e.g., subjects are being a (falsely) told that a panel of psychologists had written personality descriptions of x engineers and y lawyers and that z descriptions had been chosen from the respective pools at random) that they are expected to trust, and use. It turns out that subjects - quite reasonably do

not trust and use base rates, especially when the individuating information is at odds with the suggested base rate information. Not surprisingly, and quite reasonably, trust and use are mediated by the credibility of the source as ”an important indicator of diagnosticity and reliability” (Koehler 1996, p. 9) As we have discussed elsewhere, this is particular troublesome for psychologists who have a reputation ”for setting traps and taking delight in human failure.” (Mellers 1996, p. 1)

Catchphrase 4: EXPERIENCED BASE RATES. Directly experienced base rates, ceteris paribus, tend to be accorded more weight than indirectly experienced base rates. Experienced base rates, in other words, are considered to be high in diagnosticity and reliability. When base-rates are directly experienced through trial-by-trial outcome feedback, their impact on judgement increases, and does so quickly and dramatically (Koehler 1996, p. 6). Psychologists have hypothesized that directly experienced base rates are accorded more weight than indirectly experienced base rates because they invoke an implicit rather than an explicit learning system (Koehler 1996, pp. 6-7). Psychologists have also hypothesized that people are simply more trusting of self-generated base rates (Koehler 1996, p. 7).

Catchphrase 5: *LEARNING.* Closely related to the previous catchphrase: Tasks that involve opportunities for learning result in greater use of base rates. There is an obvious corollary here: Those tasks, ceteris paribus, that do not offer opportunities to learn can be expected to result in lesser use of base rates.

Section Three: Anomalies that have been reported in the experimental economics literature.

For the present purpose, ”anomalies” (or, as psychologists would call them, ”biases” and/or ”cognitive solutions”) are here defined as systematic deviations of subject behavior from game-theoretic predictions. One finds a great number of such anomalies in the literature (e.g., and very selectively, Thaler 199? which draws on his column in The *Journal of Economic Perspectives*; Benartzi & Thaler 1995; Benartzi & Thaler 1999; Berg, Dickhaut, & McCabe 1995; Ortmann, Fitzgerald, & Boeing *forthcoming*; Fahr & Irlenbusch 2000; Cox 2000; Fehr, Gaechter, & Kirchsteiger 1997; Fehr, Gaechter, & Kovacs 1999; Fehr, Kirchsteiger, & Kovacs 1993; Hoffman, McCabe, Shachat, & Smith (1994); Hoffman, McCabe, & Smith (1996, 1996a); Slonim & Roth (1998); Friedman 1998; or the reviews by Camerer (1995); Ledyard (1995); Kagel (1995; but see also Rothkopf & Harstad 1994; Dyer & Kagel 1996); Roth (1995); Rabin (1996); and many, many others.

We shall concentrate on games such as social dilemma/public good provision/common pool problems, gift-exchange/principal-agent game/one-side prisoner’s dilemma games, ultimatum games, trust games, moonlighting games, etc., i.e. games of conflict for which the game-theoretic predictions for one-off or finitely repeated games (with perfect information) on the one hand and indefinitely repeated games on the other, sharply differ. We shall concentrate on these games because they represent one of two prominent clusters of anomalies (the other being individual choice problems) and because more recently this cluster of games has drawn major attention of both experimentalists and theorists.

The sighting of anomalies in social dilemma/public good provision/common pool problems, gift-exchange/principal-agent/one-sided prisoner’s dilemma games, ultimatum games, trust games, moonlighting games, etc. has typically been met with one of three responses which we shall briefly discuss in the order in which they roughly dominated the research agenda:
First, robustness tests of the experimental design. The **robustness test strategy** has inevitably led to refinements of anomalies (e.g., we have learned from Hoffman et al. 1996a and Slonim and Roth 1999 that ultimatum game results seem to be quite insensitive to stakes but, as demonstrated by Hoffman et al. 1994, 1996 and Ruffle 1998, quite sensitive to entitlements and degrees of social distance) Overall, however, a significant amount of anomalies has survived these robustness tests relatively unscathed.
Second, theoretical innovations that allow experimental results to be explained within the classic paradigm. The work of the ”Gang of Four” in the early 1980's is a nice example (e.g., Kreps, Milgrom, Roberts, & Wilson 1982 on rational cooperation in finitely repeated prisoners’ dilemma games). Another nice example is the work of Stahl and his collaborators on bounded rationality and levels of reasoning (e.g. Stahl 1993, 1996, 1999; Stahl and Wilson 1994, 1995; see also Nagel 1995).
Yet another important example is Reny’s (1992) rationalization on rationality in extensive form games. Last but not least, the recent work of Goeree and Holt and their various collaborators comes to mind (e.g., Anderson, Goeree, & Holt 1998; Capra, Goeree, Gomez, & Holt 1999; Goeree & Holt 1999; and Goeree & Holt 2000, 2000a; see also McKelvey & Palfrey 1995). One of the interesting consequences of this **theoretical innovation strategy** is that anomalies become

a moving targets, for they are defined in reference to a normative model that itself may undergo updates and modifications.
Third, a number of researchers have taken anomalies at face value (e.g., and for a particularly bad example, see Rabin 1996) and have built models on ad-hoc assumptions about distributional or reciprocity preferences (e.g., Bolton & Ockenfels 2000 and Fehr & Schmidt 1999 for the former, and Rabin 1993, Dufwenberg & Kirchsteiger 1998, Falk & Fischbacher 1998, and Levine for the latter). Predictably, this **alternative assumptions strategy** has itself generated a cottage industry of experimental tests (e.g., Engelmann & Strobel 1999; Charness & Rabin 2000; and Cox 2000). An important drawback of the alternative assumptions approach and its experimental tests is the fact that documented anomalies are accepted at face value.

In the following, we will go back to the first strategy. We will first review key experimental results pertaining to the cluster of games that we delineated above and then frame the decision problems of subjects as Bayesian updating problems. Our framing explicitly acknowledges that subjects are no tabula rasa. (This in itself is not a particularly heroic assumption; see for example the noisy introspection approach of Goeree & Holt 1999, 2000a). Rather, subjects come with certain priors (or baseline beliefs) in the laboratory. There they will be confronted with individuating information ("instructions") that suggest inferences that stand in contrast to the observed and/or experienced behavior of the relevant reference group (participants in the experiment) for the particular constituent game under consideration. Quite possibly, although that's far from clear (e.g., Epstein ????), subjects will engage in (noisy) introspection. Under certain conditions, individuating information ("instructions") should be sufficient to overcome certain priors (baseline beliefs) about likely behavior in laboratory situation.

We start out with a brief description of key experimental results for each of these classes of games. Unless otherwise assumed we are talking about experimental data that resulted from test of one-off and finitely repeated games under perfect information.

Social dilemma games. Social dilemma games are here defined as 2-person two-sided prisoners' games (e.g., Ortmann & Tichy 1999 for a review of some of the literature on prisoner's dilemma games and some recent experiments; and Rapoport & Guyer 1966, Marwell & Ames 1981, Sally 1995 for important reference points). The defining characteristic is that the standard game-theoretic prediction identifies a Pareto-dominated outcome as equilibrium. This outcome is induced by (weakly) dominant strategies available to both players. The following example of a typical parameterization of a prisoner's dilemma game may serve as running example for the remainder of this paper.

Figure 1

	c	nc
c	2, 2	1, 3
nc	3,1	1, 1

In contrast, the standard game theoretic prediction - given certain strategies and discount rates/continuation probabilities - identifies the Pareto-optimal outcome {c,c} as equilibrium for the indefinite version of such games. As documented in the references listed above, social dilemma experiments persistently document deviations from the standard game-theoretic predictions for one-shot or finitely repeated games. Specifically, {c,c} action combinations typically amount to 40 to 60 percent of all outcomes. It is remarkable, and relevant to the current paper, that the experimental results have proven to be surprisingly stable to experimental manipulations (again see Ortmann & Tichy 1999 for a review of the literature and some related experiments).

Public good provision/common resource problems. Public good provision problems are also social dilemma games of sorts, only that typically they involve more than 2 people. However, the standard continuous public good provision problem (e.g., Holt & Laury 1997 and Laury & Holt 1998) can be can easily cast as a prisoner's dilemma game (e.g., to recover Figure 1 assume MPRC = 0.5, 2 subjects each can allocate one non-divisible unit to their private account or the public account, with the value of private consumption being normalized to one, and all payoffs being multiplied by a factor of 2). Likewise, prisoner's dilemma games can easily be generalized to n-person games (e.g., Frank 1988, chapter 3). As for social dilemma games, the game-theoretic prediction of outcomes of one-shot or finitely repeated games is the Pareto-dominated outcome. Likewise, the standard game theoretic prediction - given certain strategies and discount rates/continuation probabilities - identifies the Pareto-optimal outcome as equilibrium for the indefinite version of such games. As documented, for example, by Ledyard 1995; see also Holt & Laury 1997; Laury & Holt 1998; and Keser & Gardner 1999), public good/common resource experiments persistently document deviations from the standard game-theoretic predictions for one-shot or finitely repeated games, with typical contributions amounting to 40 - 60 percent of possible contributions. Once again, it is remarkable, and relevant to the current paper, that the experimental results have proven to be surprisingly stable to experimental manipulations (again see Ledyard 1995 for a review of the literature and some related experiments).

Gift-exchange/principal-agent/one-sided prisoners' dilemma games. The work of Fehr and his various collaborators (e.g.,; Fehr, Kirchsteiger, & Riedl 1993; Fehr, Gaechter, & Kirchsteiger 1997; Falk, Gaechter, & Kovacs 1999; see also Charness 2000, 2000a, and Goeree & Holt 2000) has similarly demonstrated that people do not seem to obey the prescriptions of standard game theoretic models for one-shot and finitely repeated principal-agent games as these authors find high degrees of (positive) reciprocity, i.e. "firms" offering higher wages and "workers" providing significantly more effort than predicted by standard game-theoretic models.

Ultimatum games. Closely related results for ultimatum games (e.g., Gueth, Schmittberger, & Schwarze 1982; Gueth 1995; Hoffman, McCabe, Shachat, & Smith 1994) and for that matter dictator games (e.g.,; Hoffman, McCabe, & Smith 1996) also demonstrated that people do not seem to obey the prescriptions of standard game theoretic models for one-shot principal-agent games as these authors find generous offers of proposers even in high stakes situations (e.g., Roth et al. 1991; Straub & Murnighan 1995; Hoffman, McCabe, & Smith 1996; Cameron 199?; Slonim

& Roth 1998).[Note that Slonim and Roth implement repeated game.]

Trust games. Closely related results on the trust game (Berg, Dickhaut, McCabe 1995) also demonstrated that people do not seem to obey the prescriptions of standard game theoretic models for one-shot investment games as these authors find generous investments (trust) as well as significant reciprocity on the part of those trusted. This is true even in situations where experimenters tried to exorcize trust by framing the information provided in a less favorable light and by prompting strategic reasoning on the part of the participants (Ortmann, Fitzgerald, & Boeing *forthcoming;* see also Cox 2000).

Moonlighting games. Yet another game closely related to gift-exchange/principal-agent, ultimatum, and trust games has recently been provided by Abbink, Irlenbusch, & Renner (2000). They too demonstrated that people do not seem to obey the prescriptions of standard game theoretic models for one-shot games as they find significantly more trust and reciprocity, less opportunism, and more retribution than predicted by standard models.

Section Four: ... and the environments under which they emerge.

The picture that emerges from the previous discussion suggests: (1) in games such as social dilemma or public goods provision where subjects have to make simultaneous decisions (normal form games), a significant fraction of participants ignores dominant strategies (of the affiliated one-shot or finitely repeated constituent game). (2) in games such as gift-exchange/principal-agent or ultimatum or trust or moonlighting where subjects make sequential decisions (extensive form games), a significant fraction of first-movers ignores that second-movers have (weakly) dominant actions that would put first movers at a distinct disadvantage if they would be chosen, i.e., the subgame perfect equilibrium was not selected. Table 2 summarizes.

Table 2

	Sim move	Seq move	First-mover	Second-mover
Social dilemma	**T**		n.a. (both have dominant strat)	
Public good/common res.	**T**		N.a. (all have dominant strat)	
Gift exchange/prin-ag/etc.	**T**	**T**	?	?
Ult game		**T**	dom action	dom action
Trust game		**T**	dom action	dom action
Moonlighting game		**T**	dom action	dom action

What all these games and their various experimental results is common is the fact that the ”anomalous” results that have captured experimentalists’ and theorists’ fancy were the result of one-shot (most of the ultimatum or trust or moonlighting experiments) or finitely repeated game experimental designs and implementations, i.e. settings highly unusual for the subjects.

Remark 1: The equilibrium concept that typically is employed to identify the normative solution is based on the Nash equilibrium concept which assumes that actions and beliefs are in sync and that subjects decisions are not noisy due to errors in perception, calculation, and the like.

Remark 2: Considerations such as ”When the other player thinks in the same deductive, ”rational” manner, we are very likely to end up with a ”bad” outcome. However, since the other player may think as I do (he or she is after all from the same reference group), we should both be smart enough to do the ”irrational” thing and go for the ”good” outcome. Sure there is a chance that I get screwed but what the heck do not play a role.” [Compare to Brams’ TOM.]

Remark 3: As Reny (1992) has demonstrated, given certain beliefs about the behavior of second movers the well-documented ”irrational” behavior of first movers in centipede games and similarly backward-induction requiring interactive situations may be quite reasonable. Along similar lines, Goeree and Holt have demonstrated that for important classes of games (namely those with multiple rank-based payoffs (such as those found in travellers’ dilemma, public good experiments, and coordination games) the introduction of noise in form of perceptual or computational mistakes can explain ”anomalies” in important classes of games surprisingly well. The argument is

surprisingly simple (the math less so). [Footnote here on the intuition underlying travellers' dilemma and coordination games.] More recently, Goeree and Holt have taken their argument to the extreme in models of noisy introspection (e.g., Goeree & Holt 1999 and Goeree and Holt 2000a).

Remark 4: While a significant fraction of subjects engages in actions that contradict the predictions of standard game-theoretic models, another (equally) significant fraction engages in behavior that accords with those predictions.

Section Five: Discussion

First, what the "anomalous" results of experimental tests of social dilemma/public good provision/common pool problems, gift-exchange/principal-agent/one-sided prisoners- dilemma games, ultimatum games, trust games, moonlighting games, etc. have in common, and what has captured experimentalists' and theorists' fancy, is that they were generated in one-shot (most of the ultimatum or trust or moonlighting experiments) or finitely repeated game experimental designs and implementations, i.e. settings highly unusual for most subjects who are engaged in games with related payoff structures frequently - quite probably on a daily basis - and who are likely to think little when they invoke rules of behavior that typically are appropriate for those situations (e.g., Epstein ...). In terms of the determinants of conservatism and base rate fallacy, subjects have been involved in these games and - good intuitive statisticians that they are, and having plenty of opportunity to learn what is blameworthy and what is praiseworthy (e.g., Meardon and Ortmann1996; 1996a; Smith 1982 [1759]) - they have acquired over time an intuitive understanding of the game they face (= a FREQUENTIST PROBLEM PRESENTATION based on EXPERIENCED BASE RATES/LEARNING from UNAMBIGUOUS SAMPLE SPACES). [Here possibly invoke routines literature/Smith's reasoning routines.] They have in particular experienced base rates of cooperative/reciprocal behavior. (Recall that in indefinitely repeated games the predicted outcome is close to that what has captured the attention of experimentalists and theorists alike.) Furthermore, since subjects have experienced these base rates directly, credibility problems are not an issue. Last but not least, the understanding of what is blameworthy and what is praiseworthy comes at a cost (e.g., Meardon & Ortmann 1996; 1996a; Epstein ???)

Second, attempts to explain:

We are not the first to conjecture that something funny is going on in experimental test of games where the standard game-theoretic prediction differs sharply for one-shot and/or finitely repeated games one the hand and indefinitely repeated games on the other. For example, [Hoffman et al. 1996a][Abbink et al. 2000] We conjecture that robustness tests, theoretical innovation, and alternative assumptions strategies may be missing the main point.

Other explanations we can think of are lack of credibility of the experimenter (and hence the diagnosticity of the individuating information); fear to be detected; good subject behavior; and last but not least perceptual and cognitive constraints (e.g., Cowan *forthcoming;* Goeree & Holt 1999, 2000a)..

Third, our working hypothesis is that experimental subjects come into the laboratory with "potent and reliable" base rates about social contexts that resemble those of the laboratory situation. This in turn leads to subjects not paying due attention to the individuating information about the laboratory situation (which, to make things worse, they may dismiss as non-diagnostic, non-credible, etc.) We believe that it is no coincidence that many anomalies in economics are connected to experimental designs that try to implement one-off or finitely repeated game or

decision scenarios, i.e., situations where experimental participants have to overcome the very potent and resilient base rate beliefs that they have formed in their daily routines.

In order to mine evidence on the use of base rates, let us conceptualize the decision problem that participants in experiments face as a Bayesian updating problem involving the ”typical behavior” of a reference group (participants in the particular experiment) as anticipated baseline behavior and the ‘individuating behavior” that can be inferred from the instructions for members of the same reference group. Let us take variation of a one-sided version of the (two-sided) prisoners' dilemma game of Figure 1 as example (for a derivation from primitives, see Ortmann & Colander 199?; see also Kreps 1990).

Figure 2

	nm	m
c	1, 1	0,0
nc	2,-1	0,0

Row (the agent) has choices ”c” or ”nc”; Column (the principal) has choices ”m” or ”nm”
In terms of a Bayesian updating problem, a participant will intuitively grasp (have the prior) that the baseline behavior of the players in the other role (in typical eperiments almost always members of the same microcosm that the participant is a member of) would be ”c” [Note that trigger strategies for all reasonable discount rates would induce (c,nm), i.e, the Pareto efficient outcome.]. However, the individuating information provided via the instructions suggests that in the particular set-up of a one-shot or finitely repeated game the possible behavior of other players might be ”nc”. Note that the individuating information - the data - do not really exist yet. However, what does exist is the anticipation of likely behavior of other players that may now be used as data that help to update the priors. (Goeree & Holt would call that introspection.)

We propose, in other words, that participants in game experiments do not come to experiments as tabula rasa. Rather they have, intuitive statisticians that they are (Cosmides & Tooby 1996), acquired reasoning routines for a variety of social contexts for which social dilemma/public good provision/common resource games or gift-exchange/principal-agent/one-sided reputation games, ultimatum, trust, and moonlighting games, etc. are quite representative. Over time, our participants have learned that the self-interested solution to those (indefinitely) repeated games is the kind of cooperative solution that we often see documented in experiments as ”anomalies” (but which, of course, is quite reasonable in indefinitely repeated game contexts).

Equations (3) and (4) express these thoughts in terms of equation (2) for the principal [Column] and the agent [Row], respectively:

$$(3) \qquad p(c|nc) = \frac{p(nc|c) \times p(c)}{p(\text{Instructions})} = \frac{p(nc|c) \times p(c)}{p(nc|c) \times p(c) + p(nc|\,nc) \times p(nc)}$$

$$(4) \qquad p(nm|m) = \frac{p(m|nm) \times p(nm)}{p(\text{Instructions})} = \frac{p(m|nm) \times p(nm)}{p(m|nm) \times p(nm) + p(m|\ m) \times p(m)}$$

It turns out that - given our parameterization of the one-sided prisoners' dilemma game of Figure 1, principal and agent face very different deductive problems. The agent has a (weakly) dominant action choice; the principal has a dominant choice only after anticipating that the agent will use his or her (weakly) dominant action choice. [Need to analyze this in more depth. The essence is that whatever the priors (and they reflect "c" and "nm"), the individuating information should lead to a posterior that suggests "nc" and "m" which is not what we see for many subjects.]

Conclusion:

> ”[In economics] special attention is attention is paid to the last periods of the experiment ... or to change in behavior across trials. Rarely is rejection of a theory using first-round data given much significance.’ (Camerer 1997, p. 319)

Koehler (1996) has demonstrated that the base rate fallacy maybe well be a myth and that there are plenty of other instances where ”conservatism” happens. Importantly, he has amassed convincing evidence that ”conservatism” is likely to occur when subjects (can) draw on such frequentist problem presentations from unambiguous sample spaces and that ”conservatism” is even more likely if the base rates of the problems have been directly experienced, or at least come from a credible source that can vouch for their diagnosticity and reliability.

We conjecture that the condition under which one-off or finitely repeated experimental tests of social dilemma/public good provision/common resource problems, gift exchange/principal-agent/one-sided prisoners’ dilemma games, ultimatum, trust, and moon-lighting games, etc. are typically conducted systematically favor subjects’ overweighting of base rates. Testing subjects’ rational response at measures by deductive game theory/Bayesian analysis is therefore doomed.

Unfortunately, while the thus constructed anomalies are easily to construct, they are very difficult to deconstruct because such deconstruction involves -- at least for the games that participants often encounter in their daily lives and for which, therefore, they are likely to have potent and reliable base rates -- quite possibly the construction of a counterfactual reality.

We note in closing that the controversy over the base rate fallacy is part of a major discussion in psychology between -- roughly -- proponents of the heuristics-and-biases approach and its opponents. Gigerenzer and Cosmides & Tooby and their various collaborators have demonstrated that significant parts of the biases-and-heuristics program are to some extent artifacts of the experimental design; in essence the opponents of the heuristics-and-biases approach have argued that is not with humans in those decisions but those that have constructed the laboratory situations (e.g. Gigerenzer 1991; Gigerenzer & Hoffrage 1995; Cosmides & Tooby 1996; Brase, Cosmides, & Tooby 1998). Furthermore, Gigerenzer and Cosmides & Tooby have suggested that if one tests humans in ecologically valid situations, they will - like bumblebees under ecologically valid situations - perform quite possibly differently (and maybe even according to game-theoretic predictions for one-shot and finitely repeated games).

Our hunch is that at some point in the hopefully not too distant future experimentalists will likewise come to realize that the results of experimental tests of one-shot and finitely repeated social dilemma/public good provision/common resource problems, gift exchange/principal-agent/one-sided prisoners’ dilemma games, ultimatum, trust, and moon-lighting games will likewise come to realize that the problem in those decisions was not with human performance but with the laboratory situation.

[Issue of great importance because it affects the incentive compatible design of all institutions and organziations.]

Appendix A

Example 1. Holt & Anderson (1996, pp. 179/80; for similar examples see Eddy 1982; Hoffrage & Gigerenzer 1995; Gigerenzer 1996) tell the (true) story of a man who was told, following a first-stage test, that he had the virus that caused AIDS, and who committed suicide before follow-up examinations. Given a base rate p(H) of 0.4% (”about one in 250 at that time”), a hit rate p(D/H) of 100% (perfect), and a false positive p(D/**1**H) of 4%, equation (1a) tells us that **9** = p(H/D) = p(D/H) x p(H)/ [p(D/H) x p(H) + (D/**1**H) x p(**1**H)] = 100% x 0.4%/[100% x 0.4% + 4% x 99.6%] = 1 x 0.004/[1 x 0.004 + 0.04 x 0.996] = 0.09, or less than 10 %.

Example 2. Gardner (1995) illustrates how a law firm that just hired a new lawyer named Kane might go about assessing her potential. Kane has yet to try a case but the firm knows from experience that two kinds of lawyers survive its screening process: ”Stars” and ”ordinary” ones. Star lawyers win 75% of their cases, ordinary ones win 50% of their cases. The firm also knows from experience that only 10% of its newly recruited lawyers turn out to be stars. Preferably that’s the kind of lawyer that the firm wants to give a long-term contract to. So, how does the law firm figure out whether Kane is a lawyer it wants to keep? Or, in other words, what is the posterior probability of Kane being a ”star” once she has won her first trial? Given a base rate p(Star) of 20%, a hit rate p(win/Star) of 75% (good but imperfect), and a false alarm rate p(win/Ordinary) of 50%, equation (1a) tells us that **9** = p(Star/win) = p(win/Star) x p(Star)/ [p(win/Star) x p(Star) + (win/Ordinary) x p(Ordinary)] = 75% x 10%/[75% x 10% + 50% x 90%] = .75 x 0.1/[.75 x 0.1 + .50 x 0.9] = 0.1429. If Kane wins her second trial too, the posterior probability of Kane approaches 20%.