A Gricean analysis of understanding in economic experiments

Jones, Martin

Publication date:
2004

Dundee Discussion Papers in Economics

A Gricean Analysis of Understanding in Economic Experiments

Martin Jones

Department of Economic Studies, University of Dundee, Dundee. DD1 4HN

Working Paper
No. 171
July 2004
ISSN:1473-236X

A Gricean Analysis of Understanding in Economic Experiments

Martin K. Jones[1]

Abstract: There are many different ideas of understanding in the experimental economics literature, leading to much debate and conflict between various methodological positions. It is argued that a Gricean approach to the notion of understanding is one which makes most sense in the context of experimentation and provides a robust theoretical background for assessing understanding in practical experiments.

JEL Classification: B41
Keywords: Grice, Experiments, Understanding

[1] Department of Economic Studies, University of Dundee, Dundee. DD1 4HN

Introduction

The introduction of the experimental method into economics has resulted in a revolution in the techniques used and the knowledge gained in the discipline. Together with the growth in experimental methods there has been a gradual growth in the methodological discussion of experiments and their role in economics. Some of this discussion has come from experimentalists themselves either laying down "methodological norms" for the field (e.g. Smith 1982) or discussing problems within the field (e.g. Starmer 1999).

This methodological discussion has become more interesting in recent years with the emergence of a variety of "styles" of experimentation and the methodological prescriptions which the proponents of these different styles advocate (see the contrast between Starmer, Binmore and Loewenstein's articles in their 1999 Economic Journal Symposium). In addition to the discussion within economics there are closely related discussions in psychology which are relevant for economic experimentation (See Tversky and Kahneman (1996) and Gigerenzer (1996) for an example).

For the purposes of this paper I would like to discuss and critique a collection of closely related arguments about experimentation which seem to have gained widespread currency amongst many experimentalists. These arguments share an underlying assumption which concerns the role of *understanding* in an experimental situation. Put simply, it is assumed that, at least in certain experimental situations, there is a systematic gap between the subjects' understanding of the experiment and the understanding of the experimenter. The systematic nature of this gap leads to radical conclusions about the correct interpretation of results and the design of experiments.

This requires explanation as it is easy to confuse issues in the discussion ahead. Here we will identify a person's "understanding" of an experiment with that person's *interpretation* of what the experiment's instructions, format, goals, payoffs and the general layout of the experiment mean given a certain level of background knowledge. It is assumed that both the experimenter and the subject have their own "understandings" of the experiment. We will not bother to question whether there is an absolutely "correct" interpretation of this knowledge, but will instead focus on whether these interpretations can be brought into agreement.

Principally, we will be looking at whether subjects can appreciate the meaning of the information communicated by the experimenter although we will also look at the situation the other way round. In the former case, the experimenter's understanding is the criterion by which we shall judge whether the subject has made a mistake or "misunderstands" the experiment or not. So, for example, a "full understanding" would come about when the subject' understanding is fully aligned with that of the experimenter. Similarly a subject "understands" the experiment if she has a similar alignment. Another way of looking at this is to see the aim of the paper as trying to define and analyse *intersubjective* understanding between the experimenter and subject rather than objective understanding of an absolute meaning.

"Understanding" in this context simply means being able to know the meaning of the *content* of a given problem in an experiment. It does not include the psychological factors which go into making a decision since these are observed as part of the *response* of the subject rather than part of their understanding of the experiment. These psychological factors have a source which is assumed to be distinct from their perception of the experiment. This means that the common notion of "self-understanding" where people come to terms with their own emotional reactions is not included in this definition. Understanding therefore is the interpretation of transferred knowledge from the experimenter to the subject.

If it is assumed that there is a systematic gap between understandings then this can lead to a variety of conclusions depending on how this is interpreted. The two interpretations discussed here are as follows:

a) The subject does not understand the experiment because of a mistake[2] (for example identifying incorrect social norms (c.f. Binmore 1999)), some perceptual bias (Plott 1996) or because the experiment has been presented in the incorrect format (c.f. Gigerenzer 1996)). It follows that, while the experiment in its original form may lead to normatively incorrect behaviour, this can be corrected by allowing the subjects to find out their mistakes for themselves or by putting the experiment in a different format.

[2] In other words the subject's understanding has deviated from that of the experimenter.

b) The experimenter systematically fails to understand the subjects' interpretation of the experiment. It follows that an interpretation of the subjects' behaviour as normatively incorrect is misconceived. It may simply be the case that the subject is normatively correct within his or her own interpretation (see Plott 1996).

It can be seen that these two interpretations have a lot in common. In fact they are basically two different ways of saying the same thing except that the location of the "error" differs. In (a) it is the subject who is seen as making the error, while in (b) it is the experimenter. In both cases there is a belief that the lack of understanding (on the part of the experimenter or the subject) is caused by a fundamental flaw in the experimental design. There also seems to be an assumption that subjects are usually normatively rational[3] even though an experimenter's and subject's understandings may not coincide. Finally, it is assumed that only particular types of experiment (if any) are "legitimate" as tests of rationality. This effectively blocks off a large number of experiments which have been used in the past to demonstrate irrationality (such as those used to empirically verify the Allais paradox).

There is an inevitable mixing of normative and empirical considerations in this analysis. This follows from the basic aims of many experiments in economics: to test for normatively correct decision- making against the evidence. For example, it may be decided that the experimenter's interpretation of what is normatively correct in an experiment is too "tight" in the sense that it excludes many actions which could be seen as normatively justifiable. The normative standards could then be adjusted to compensate by allowing more actions to become acceptable. This would then change the empirical results in the sense that more actions would become normatively correct.

This normative aspect means that our analysis will stretch beyond the strict confines of experimentation to discuss similar issues in decision theory. Broome (1991), for example, has made similar arguments to (b) from a normative point of view even though he does not explicitly talk about experiments. However, his arguments could, in principle, be used as a critique of experimental results in that he proposes a normative criterion for observed decisions which is far looser than that

[3] This contrasts with the conventional interpretations of "Irrational" results which is that subjects are violating normative standards.

usually accepted by economists. This would eliminate many refutations of expected utility based on conventional criteria. This is discussed in detail later on in this paper.

It should be noted that the issue of understanding has not been discussed in any great depth in the economics literature. Understanding, when it is introduced, often has an *ad hoc* role or is often not mentioned at all. In spite of this, there are a variety of possible criteria for understanding which exist either implicitly or explicitly in the literature. However they are not analysed in any great detail and the implications are not formally drawn out. Some of these criteria will be discussed in this paper. Having discussed these, it will be argued that arguments about understanding in experiments can be put into a more systematic philosophical framework based on the writings of Grice (1989). In that light, some of these criteria will be examined and compared with the Gricean model.

During much of this paper the discussion will centre around experiments in either game theory or decision theory. There are many reasons for this. One is that this is the area where normative theory (in the form of Expected Utility) has been debated most. Another reason is that decision theory has many of the most clear- cut results in the literature in which there are regular violations of normative theory. It also has the distinction of being the area of experimental economics which has been most discussed in the methodological literature.

The paper will be split into seven sections. Sections 2 and 3 will look at the problems outlines above in more detail. Section 4 will introduce the relevant part of Grice's work as well as deriving useful criteria while section 5 will apply these to the problems in sections 2 and 3. Section 6 compares these claims with other applications of Grices' ideas in an experimental context. Section 7 concludes.

2. The Role of Understanding in Experimental Methodology

There has been considerable interest *in practice* in the problem of understanding amongst experimenters. It would be reasonable to say that the general stance of the experimental practitioner towards encouraging the subject to a full understanding of the experiment is fairly pragmatic. In general, there is a set of well-thought out instructions to explain the experiment, usually in conjunction with illustrations of the points made. Many experiments usually have "practice" runs where the subjects can try out the mechanics of the experiment. Some also have mini-

questionnaires to test understanding of important parts of the experiment. All experiments tend to be piloted in sessions where experimenters can ask subjects informal questions about the experiment and improve the experiments based on these answers. In addition all possible effort is usually made to make the experiment as clear and transparent as possible.

These explicit attempts to ensure understanding are supplemented by other aspects of the experiment[4]. All experiments are incentivised with real and explicit incentives where the method of achieving these incentives is obvious to the subject, overrides any subjective costs of being in the experiment and is increasing in "good" outcomes (Smith 1982). It is an implicit assumption that these incentives will help the subject concentrate on the experiment and so more easily understand it. A further requirement, particularly in economic experiments, is that there should be no deception in the experiment. It is thought that deception lowers trust in the experimenter and, it could be argued, prevents subjects from understanding the purpose of the experiment. Deception could also increase misunderstanding by encouraging subjects to try to "second- guess" the experimenter and guessing wrong.

It will be appreciated from this that the testing of understanding in economic experiments amongst practitioners is more of a craft than a science. Many of these aspects are purely pragmatic- the use of examples and testing, for example, could be seen as an educational technique applied to experimental research. Some are based on ad hoc "feelings" about how to make subjects understand. Others are simply aspects of good experimental design which, incidentally, lend themselves to understanding.

Much of the discussion of understanding in experiments in the more methodological literature is based around categorizing various types of error and correcting them. Starmer (1999), for example examines Smith's axioms for a well designed experiment (for example the salience of incentives) and then suggests solutions for violations of these axioms (Simple designs, greater transparency, familiarisation with the setting of the experiment and testing knowledge). Binmore (1999) stresses that people should not be put in too complicated a situation, the problem should be reasonably simple, incentives should be adequate and there should be room for trial- and error learning.

[4] This and the following remarks tend to apply more to economics experiments than to psychology experiments.

There is no *theory* of understanding within the experimental economics literature apart from the generally accepted observation that, if subjects consistently answer experimental questions "correctly" (i.e. in line with Expected Utility Theory) then they must understand the experiment[5]. Logically, it does not follow that if subjects do not follow EUT then they do not understand the experiment. There are other possibilities, including the possibility that subjects do understand the experiment but have a different view of what is normatively correct (they could, for example, be a follower of Allais (1953)).

One theory which has gained wide acceptance in recent years and is closely related to a theory of understanding is the Discovered Preference Hypothesis (Plott 1996, Binmore 1996). A central part of this is the belief that the normative force of Expected Utility is such that all reasonable people would want to follow it and in fact do so. It follows that if subjects do not seem to outside observers to follow EUT in experiments then this must be the result of one of a series of possible subject errors relating to the content or information provided in that particular experiment. Understanding, or a lack of it, forms a large part of the reason for why subjects make these errors (see Cubitt et al. 2001).

This is a radical proposal in that it assumes that Expected Utility is the only possible normative standard and that subjects would conform to it rather than any other standard[6]. Furthermore (see Plott 1996) the question of whether subjects do or do not have consistent preferences is effectively put beyond testing. It is *assumed* that subjects have consistent preferences and that errors are simply the result of not having sufficient knowledge or experience of the particular experiment. Observed errors under this view are simply the result of consistent preferences over erroneous "frames" of the experiment.

This hypothesis is, amongst other things, a critique of the designs of certain experiments. If these experiments are "badly designed" then people will not be observed to maximise utility. In order to correct the design of such an experiment then it will have to incorporate corrective principles such as transparency, sufficient incentives and learning opportunities. Experiments are "transparent" if unnecessary

[5] However, even this does not always follow. As Cubitt et al. (2001) point out one can arrive at observed EU maximising even with inconsistent preferences through a process of contamination in learning.

[6] This is a dubious assumption from a normative point of view (see Loomes and Sugden (1982) on Regret Theory as a normative standard) but also has been heavily criticised in the literature (see Kahneman's (1996) reply to Plott for an example).

complications are stripped out and the subject is given sufficient relevant information to carry out the task. Incentives have their usual role in experiments of concentrating subjects' minds while learning involves repetition in order to familiarise the subjects with the tasks.

In essence, the view of understanding taken by the advocates of discovered preference is simple. Assume *ceteris paribus* that subjects make none of the other possible types of error apart from errors in understanding[7]. In this case a subject can only be said to understand an experiment if they are observed to maximise Expected Utility. If a subject is not observed maximising Expected Utility then, according to this theory, (although as stated above this does not *logically* follow) the subject does not understand a question. It is possible for subjects to *systematically misunderstand* experiments because of the existence of social norms or other framing effects. This means that it may be that some aspect of the experiment reminds subjects of some norm existing outside the experimental laboratory (or alternatively may cause them to "frame" the problem in a different way) and a systematic error may occur as a result of following this norm.

The aspect of the Discovered Preference hypothesis which has been criticised most heavily in the literature is that of repetition. According to Binmore (1999), the subject may start out making an error, possibly a systematic error. For this reason a single game is useless because the subjects do not understand the experiment fully and the effects of external "norms" have not been stripped out. The game is only valid once it has been repeated several times since, in this way, mistakes will be discovered by the subjects and eliminated on future repetitions. This process of "trial and error" learning will result in the "true" preferences being discovered after several repetitions. This has the effect of reducing "contextual" effects and so increasing the external validity of the experiment.

Tversky (1996) has criticised this on the grounds that this does not represent learning in any meaningful sense since the feedback in such experiments is far richer than in the real world environment. To say that such experiments increase external validity is peculiar, given how few situations resemble such a narrowly defined repetitive situation (see also Starmer 1999). Loewenstein (1999) has also made criticisms along these lines, questioning how representative such a situation could

[7] Plott seems to assume that most errors are errors in understanding.

ever be; how could the sixth repetition be more valid than the first? There is always the risk of confounding factors, such as boredom, as a result of repetition which may influence certain experiments (such as the ultimatum game[8].)

Cubitt et al.'s critique extends this by noting that repetition increases the complexity of a game and thus makes it far more difficult to understand. If the instructions for one game are hard to understand then adding on repetitions gives an extra layer of complexity which needs to be understood. Furthermore, repetition increases the risk of contamination either from a player's previous play or from the play of other people. This increases the possibility of error on the part of the subject or may even create a "false positive" of subjects who are not Expected utility maximizers but who choose in a way consistent with it.

In response to this Cubitt et al. propose that in place of repetition an experimenter should concentrate on *control* of experiments. If subjects do not understand an experiment then it must be a failure of control within the experiment. In general one should follow the precept that subjects do understand an experiment (given that the experiment is simple, instructions are clear etc.) unless one is specifically testing for a lack of understanding. If such a test is performed and some misunderstanding comes to light then a new control should be introduced to prevent this misunderstanding.

A variant of the "Discovered Preference" hypothesis can be found in the psychology literature (see Gigerenzer 1996 and Kahneman and Tversky's 1996 reply). According to Gigerenzer much of the work done by psychologists on the heuristics and biases programme, as exemplified by Kahneman and Tversky's work, is misplaced. Many of the so- called biases from normative standards in fact disappear when the questions are presented in a "suitable" format. The example given is Kahneman and Tversky's "conjunction" problem[9]. Gigerenzer has shown that, when done in the original probability format, the conjunction bias holds but it disappears if the problem is redescribed in a frequency format. Gigerenzer interprets this as showing that subjects cannot be demonstrated to be irrational by Bayesian standards. However, unlike the advocates of Discovered Preference, he does not seem to believe that Bayesian EUT is the only possible normative standard and instead critiques this

[8] As Loewenstein points out- how many times can one be outraged by an unfair division of money?
[9] Also known as the "Linda" experiment. Subjects are given a description of "Linda" and are asked the likelihood that she is either (a) a bank teller or (b) a bank teller and a feminist. A substantial proportion of people choose (b) violating the axioms of probability (since (b) is a subset of (a)).

view, pointing out that Bayesianism is controversial in the statistics literature. In spite of this dislike of Bayesianism as a normative standard, Gigerenzer's argument, like that of Plott and Binmore, does suggest that the reason for subjects not to obey Bayesian rules is because they have "cognitive illusions" as a result of having information presented in the "wrong" format. For this reason the two points of view can be seen to be similar.

It can be seen that there are two different views of understanding at work in this debate. On one hand, Binmore and Plott very closely link understanding to expected utility. Roughly, being observed maximising expected utility in an experiment (holding other errors constant) is the criterion for understanding. If one does not maximise expected utility then, excluding other types of subject error, it must be because subjects do not understand the experiment. Gigerenzer seems to have a looser position than Binmore and Plott in that a subject can be brought to obey a certain normative standard (not necessarily EUT) by changing the format of the experiment. In this sense they can be said to "understand" the experiment if they have a "suitable" format.

Cubitt et al and Loewenstein, by contrast, have a pragmatic view of understanding. Sources of misunderstanding need to be rooted out empirically and, if found, need to be controlled. A subject is assumed to understand an experiment if it is properly controlled and the framework is properly tested. Understanding in this framework is not attached to one's performance compared to a normative standard but is assessed by direct testing and an empirical view of understanding. This view is essentially *ad hoc* in that there is no explicit theory of understanding underlying the reasoning.

3. The Role of Understanding in the Analysis of Decision Theory

The implicit disagreement about the definition of understanding in the experimental methodological literature is paralleled by another implicit disagreement in the decision theoretical literature. One part of the philosophical discussion about rationality in economics has focussed on the normative consequences of violations of Expected Utility. As was discussed above, this becomes important in the experimental field because of the interpretation, in terms of rationality, given to results in

experiments. If too loose an interpretation is given then this will mean that experimental tests become virtually useless as tests of "rational" behaviour.

In this section it may be useful to see the analysis in experimental terms. Since the two authors discussed in this section, Anand (1993) and Broome (1991), are discussing normative considerations which apply to all decisions this means that these considerations also apply to decisions made in experiments. The normative analysis can be seen as mainly applying to the subject in the experiment while the experimenter may be seen as an outside observer who sees normatively correct or incorrect choices without knowing the subject' understanding of the experimental problems. The line of argument discussed by Broome suggests that the experimenter may be taking too constricted a line on what constitutes rational behaviour if she declares observed Expected Utility violations to be irrational.

While there has been a long tradition in economics of regarding expected utility as the only possible type of rational behaviour there have been alternatives proposed. Indeed Allais' (1953) original purpose in presenting his paradox was to demonstrate the *normative* point that the independence axiom was not intuitive[10]. Since then there have been a variety of claims for the rationality of various different theories. One example of this is Loomes and Sugden (1982) who defend Regret Theory by pointing out that it is perfectly reasonable, if one experiences regret, to include this emotion in one's value function. This means that this value function may then violate the independence axiom[11].

However, it has been argued by Broome (1991), Anand (1993) and Tversky[12] (1975) that looking at normative or descriptive observed violations of expected utility as violations of rationality depends upon the consequences of a particular choice problem being defined in the same way by the theorist and by the agent. Define the term "non- separating" as referring to those factors, such as regret, which cause one consequence to be influenced by another consequence. Suppose that there are some factors in a choice problem which are non- separating. If the result of this is that agents preferences are not consistent with the independence axiom then it is possible

[10] Allais was using experimentation in the unusual sense of testing the normative intuitions of his subjects, most of whom were experts in the field of choice theory.
[11] In fact Regret Theory can lead to violations of transitivity as well.
[12] Tversky here is commenting on normative aspects of rationality rather than good experimental design as when he criticised Binmore. This accounts for the seeming inconsistency of his point of view.

for the agent to *respecify* the consequences to incorporate these factors and so restore consistency.

As an example look at the Allais paradox (Allais 1953):

Lottery 1: Certainty of receiving £10

Lottery 2: 0.05 chance of receiving £20
0.9 chance of receiving £10
0.05 chance of receiving £0

Suppose (as is often the case) that lottery 1 is preferred to lottery 2.
Subjects then have to choose between two further lotteries:

Lottery 3: 0.1 chance of receiving £10
0.9 chance of receiving £0

Lottery 4: 0.05 chance of receiving £20
0.95 chance of receiving £0

The independence axiom states that, if one follows expected utility it is necessary once one has chosen lottery 1 over lottery 2 to choose lottery 3 over lottery 4. However, there is a tendency amongst many people to choose lottery 4 over lottery 3 instead. This violates the independence axiom and, to the outside observer, suggests that the observer is not maximizing Expected Utility.

Suppose that we respecify the consequences of the lotteries and define lottery 1 as:
10% of £10 and 90% of £10
This corresponds with lottery 3:
10% of £10 and 90% of £0.

In these two cases we could respecify the 10% state as two different states e.g.:

Lottery 1:10% of £10 when the alternative is £10; 90% of £10
Lottery 3: 10% of £10 when the alternative is £0; 90% of £0.

This would create new lotteries 1 and 3 for which it would be perfectly reasonable to prefer lottery 1 to 2 and lottery 4 to 3. In this way expected utility has not been violated since the two 10% states are not the same and so there is no reason for them to be treated the same. In this way a simple change in the form of words leads to independence being restored and the agent playing these lotteries would be an Expected Utility maximizer.

The respecification outlined above is what we will refer to as a *linguistic* respecification since it simply reorganises the information available. This differs from Broome since much of his argument is actually based around examples where respecification involves incorporating psychological variables into the payoffs such as Regret or Disappointment. As we mentioned earlier, we will ignore these variables since they do not come within our definition of understanding. However, Broome's argument applies equally well to linguistic as well as to behavioural respecification and so the argument will be on the former rather than the latter. From this point onwards when "respecification" is mentioned then it will be linguistic rather than behavioural.

It will be argued here that the process of linguistic respecification outlined above is a reinterpretation of the agent's understanding of the lottery. In their original forms the lotteries correspond to what the experimenter understands. However, once respecified then the subject's understanding has effectively deviated from that of the experimenter and their "understandings" can be said to be different[13]. In terms of interpretation (b) in the introduction, the agent now has a "correct" understanding of the lottery.

However, as has been pointed out by Sen (1985), Machina (1981) and indeed Anand and Broome, this respecification story is not sufficient to support the normative (or descriptive) status of Expected Utility Theory simply because it lacks content. If any set of lotteries can be respecified to support Expected Utility Theory then it follows that none can be excluded and so Expected Utility loses all normative force. Respecification needs to be restricted in order for it to be useful.

Broome's solution to this problem is through an idea called the "rational individuation of justifiers". This determines whether two states can be distinguished as separate or whether they should be merged into one state. Two states can be said to

[13] Since we are looking at normative considerations here, we do not have to look at other types of subject error.

be different if it is rational to have a preference between them[14]. This allows for a subject to determine whether it is reasonable or not to differentiate between two states. It also has the advantage that it gets around the problem of lack of content in that some states (those where agents are indifferent) cannot be distinguished. Broome claims that, given the rational individuation of justifiers, the independence axiom has content but there are no counter examples.

Broome's principle has been severely criticised in the literature. One criticism (by Temkin 1994) is that it is difficult to think of a situation where some rational preference in Broome's sense could not be invented. Temkin points out that in situations where we commonly accept complementarities (for example having red wine with red meat), we can most easily distinguish states (so the state "red wine and beef" is commonly seen as different from "White wine and beef"). In this case it is certainly rational to have a preference between the two states. However, if this is true then one must dispute whether Broome's principle really does have any content. If complementarities *always* result in a rational preference once consequences are respecified then the principle is empty of content.

Another criticism, briefly mentioned by Hausman (1993), is to question the scope of Broome's principle. If the principle works in the area of decision theory then it should also work in other areas of economics such as consumer theory. In this case there are certainly complementarities between goods since consumer theory assumes that goods are complements or substitutes for each other. While these goods could (or indeed should according to Broome's principle) be redefined to remove these complementarities this would be an unnecessary complication to the theory of consumer choice.

Anand (1993) has put forward a different analysis of the respecification of states. He concentrates on the transitivity axiom rather than the independence axiom[15]. Anand argues that it is always possible to respecify states so that transitivity is not violated and expresses this in terms of a "translation theorem" which states that all transitive behaviour can be redescribed as intransitive behaviour and vice versa.

Transitivity is presented as an essentially linguistic phenomenon. Linguistic conventions are used to assess whether choice behaviour is transitive or not. When

[14] Note that this does not say that one does have a preference- merely that it would be rational to have one. "Rational" is not well- defined here (it cannot be expected utility since this would lead to circular reasoning) but the emphasis is on "preference" rather than "indifference".
[15] Although he states that his analysis is more general

new conventions are specified then this affects one's judgement on the transitivity of choices. These conventions are essentially arbitrary and so without agreement on these conventions there is no possibility of finding out whether an agent's behaviour obeys rational choice or not. Anand draws the implication that for axioms to have content they must be accompanied by rules for constructing the states over which the axioms are defined i.e. there must be agreement on the linguistic conventions. For the axioms to have *normative* content, this means that these linguistic conventions would need to be conventions of a "rational language". Anand claims that the idea of such a language is incoherent.

Anand derives the notion of a "rational language" from Hirsch (1988). In his paper Hirsch tries to formulate a set of rules which would create a "good" language - namely one that would exclude "strange" concepts such as "grue"- green before a certain date and blue after - or "gricular" , green and circular. Hirsch looked at a broad class of possible rules, covering metaphysical rules, epistemological rules, rules for explanatory power, pragmatic rules and rules for learning ability and found that there were no plausible rules in any of these categories which excluded strange concepts such as "grue" etc. Anand believes that a "rational language" is one which can exclude all such strange concepts.

Anand concludes that the only way in which we can determine whether a subject is transitive or not in descriptive Expected Utility Theory is with respect to a "specified language" which has been given *a priori*. Anand's conclusion seems to be negative- it is not really possible to have a transitivity axiom with normative content, while for it to have content at all requires agreement on linguistic conventions.

However, it is not clear why Anand believes that a "rational language" in this particular sense is essential for transitivity to have normative content. All that Hirsch shows is that there are no rules to exclude "strange concepts" from a language. In principle, of course, this would mean that there is no reason to believe that one language rather than another is better when attempting to linguistically divide the world "at the joints". However, it is normatively unclear why one language rather than another would make a "better" or "more rational" base for decision theory. Hirsch's results do not give us a basis for declaring a language "rational" or "irrational". Instead it would seem that the best language for a rational person to speak would be one in which they would be understood by other people. Rationality would then be

built on a given language rather than trying to find a rational language to go with rational actions[16].

Given these two viewpoints on normative theory it is interesting to ask how the concept of understanding fits in to these frameworks. We have defined a *linguistic* respecification as a change in the subject's understanding of the experiment so we can see that Broome's principle, when applied to this type of respecification, simply acts as a constraint on the number of different subject understandings which are permissible without becoming irrational. In general this is quite broad as Broome's principle would accept that a large number of respecifications would be rational. Normatively, there is no reason for any more than a loose restriction on the number of understandings permissible and it may be reasonable for a subject in an experiment not to have the same understanding as the experimenter as long as they maximise utility with this understanding.

Anand's views are more extreme than Broome's and lead to a pessimistic view of understanding. Unless we have the subjects' representations of the problem then we cannot tell whether their understanding corresponds with the experimenter or not. In particular, there is no a priori criterion which would govern whether a person's understanding is or isn't reasonable. It follows that there is no normative restrictions on the type of understanding held by the subject and that there is no particular reason to have one understanding over another. On the other hand, Anand's views on subject representations do include a possible solution for the *descriptive* problem, namely agreed linguistic conventions. It is to these that we shall return in the next section.

4. Another criterion for understanding: Gricean ideas on meaning

So far we have looked at the subject of understanding in economic decision making and experiments. The subject is not easy to analyse because, quite often, the idea of understanding is conflated with other notions. In general we have found that there are a variety of possible criteria for understanding, ranging from Anand's notion that it is virtually impossible to have a normative criterion of understanding, through Cubitt et al.'s pragmatic view of understanding to the ideas of "Discovered

[16] Indeed, given that a "rational" language in Anand's terms would not be our own and would not be generally known, it must be questioned how rational it would be to *use* such a language.

Preference" where only if one maximises utility can one be said to understand a problem.

In general these views fit in with the notion that there are two types of understanding; the experimenter's and the subject's and that either of them could be mistaken. The subject could be mistaken as to the "proper" normative standard or the layout of the task or the experimenter may be mistaken as to the understanding of the subject. In both cases there is a divergence of understanding between the subject and the experimenter. However, one flaw with these views is to see understanding as something which both the experimenter and the subject carry out independently of each other. In reality, understanding is often intimately linked to the idea of conversation or (more generally) communication. It requires both the person communicating and the person receiving the communication in order for the latter to understand.

This is particularly true in an experiment where the experimenter communicates instructions and the subject receives them. It may seem strange to model an experiment as a conversation but, from the point of view of this analysis, this is perfectly reasonable since this conversation does not have to be two sided. Only one of the two participants needs to be communicating information while the other listens. Furthermore, each subject in the experiment can be construed as having a conversation with the experimenter.

This model of understanding as being linked to communication also allows an analysis of meaning. Understanding is intimately linked to meaning in that when one understands the words in a sentence then one also knows the meaning of the sentence. Therefore any theory of understanding will also implicitly suggest a theory of meaning. This means that it is plausible to look for a theory of meaning when one is looking for a criterion for understanding.

The theory of meaning which takes advantage of the communication aspect of experiments is that put forward by Grice (1989)[17] and enhanced by Schiffer (1972) and Avramides (1989). This theory of meaning can easily be linked up with the idea of understanding because of its use of the intentions and responses of the receiver of the communication in order to define meaning. Principally it conceptualises meaning

[17] Grice's theory of meaning has been used as a tool of analysis in experiments before- these ideas will be discussed later in the paper.

as being defined by its use in communication and the responses of the listener to that communication.

Grice's work on meaning aims to give an *analysis* of meaning or, in other words, to find the criteria for the application of the concept of meaning[18]. Grice uses a series of preliminary assumptions about speakers and listeners in a conversation (e.g. listeners know that speakers may make mistakes or that speakers know the usual conventions of language) in order to construct his theory. These are "givens" which prepare the ground for the main part of the theory. In the following analysis it is assumed that these conditions are fulfilled in an experimental setting.

Grice's analysis is based around the notion of an "utterance". This term is used because his theory of meaning covers more than just spoken words. It can include all sorts of communication including written words and even symbols or pictures. The term "utterance" is a general word which encompasses all of these ideas. Grice divides the meaning of an utterance up into several categories. The two which we will discuss here are "speaker meaning" and "timeless meaning". Speaker meaning is the meaning of the utterance at a particular point in time, allowing for contextual and other non- linguistic factors. Timeless meaning is that part of speaker meaning which is constant across utterances- usually the linguistic element.

Speaker meaning was the focus of Grice's main analysis which looked for the conditions for which this type of meaning can be necessarily and sufficiently defined. Speaker meaning was seen as being basic to the analysis and timeless meaning flowed from it. The reason for this is simply because communication consists of far more than the timeless meaning of an utterance and so to have full analysis requires one to look at the non- linguistic as well as linguistic elements of that utterance. Grice (followed by Schiffer and others see Avramides(1989)) proceeded by use of a series of definitions which gave the definition of meaning in terms of the speaker's and listener's knowledge, intentions, beliefs and expectations. We will not go through the entire process of definition as it is a complex and subtle argument which is not relevant for the analysis here. However it is sufficient to mention that this defines speaker meaning for any utterance of a speaker.

Grice went on from his definition of speaker meaning to a definition of timeless (i.e. language) meaning. In general, timeless meaning is defined by Grice

[18] Specifically Grice is looking at "non- natural" meaning i.e. when we say that x means y we are not stating that x is caused by y (as in natural meaning) but that y explains x in some way.

according to the knowledge of the group in which a speaker is socially situated. If other members of this group have a certain utterance in their repertoire of utterances then this will cause the speaker to retain the utterance in her own repertoire of utterances when trying to convey a similar concept. Furthermore, for timeless meaning to function within a language then an utterance must be dependent for its occurrence in conversation on non- natural features or a convention- natural features depend too much on context to be "timeless". A language is structured from a series of conventions governing utterances in conversation.

Grice never really gave a fully coherent account of timeless meaning so the "Gricean" account of timeless meaning which has emerged as the most popular is that of Schiffer (1972) (See Avramides 1989). Schiffer related timeless meaning to the idea of a convention as defined by Lewis (1969). In this view a convention is seen as a coordination game[19] between the speaker and the listener in a conversation. When they manage to coordinate on a meaning for an utterance then this meaning, after repetition of this coordination over a period of time, would become a convention and so part of the language. This idea of the timeless meaning of a word as a convention allows for language to be seen as an autonomous entity. A sentence may be uttered and the words have meaning even if that meaning is different from the meaning under the Gricean account of speaker meaning.

In fact it is an essential part of Grice's account of meaning that there may be a divergence between the timeless meaning of an utterance and the speaker meaning. It is this divergence which allows for what is known as "conversational implicature". A conversational implicature emerges simply because an utterance by a speaker may have meaning beyond the formal meaning of the words (i.e. timeless meaning).

Grice assumed that one could analyse conversational implicature using the idea of a conversation being, by its very nature, a cooperative affair[20]. Each participant in a conversation has a common purpose or direction with all the other participants. Grice's "cooperative principle" states that one should make a conversational contribution according to the accepted purpose or direction of the conversational exchange. This cooperative principle effectively excludes some moves in a conversation from the start (e.g. always talking nonsense).

[19] Strictly, Schiffer did not see coordination as the motive but rather a desire to understand.
[20] This does not exclude the possibility of lying. However lying is assumed to be a comparatively rare event. Otherwise communication would lose its rationale (if everyone lies then there is no point in communicating to receive false information).

From this cooperative principle Grice derived *conversational maxims* which would ensure the maintenance of the cooperative principle. There are four maxims, those of Quantity, Quality, Relation and Manner. The Quantity maxim is the requirement to make one's contribution as informative as possible but not more informative than required. The Quality maxim is the requirement to make what one says as truthful as possible without saying false things or things for which one does not have evidence. The Relation maxim is simply the requirement that one should be relevant. The Manner maxim relates to how one expresses information. So, for example, one should avoid ambiguous or absurd expressions and be brief and orderly.

These four maxims are the most general possible maxims and apply to most types of communication[21] but they are not the only maxims possible since other more context specific maxims apply to particular situations. Also these four maxims are too general because more detailed statements are needed for applications. They need not hold all the time for all conversations but if they are violated either by mistake or on purpose then this will change the meaning of the words beyond that given by the timeless meaning of those words.

As has been mentioned before, this can help us to construct a theory of understanding. With Grice's theory of meaning this is comparatively easy since the theory is structured in terms of the intentions and knowledge of the person who is listening to the speaker. Therefore, if a listener has the reactions which are necessary for a certain utterance to have a given meaning then this means that the listener *understands* the utterance to have that meaning. It follows that the conversational maxims are necessary criteria for the understanding of the listener as well as defining the meaning of an utterance.

Grice viewed communication as a special case of purposive rational behaviour. People talk in order to achieve the goals of the conversation. They generally obey the conversational maxims because this means that conversational exchanges will be profitable. If people do not obey the maxims then meaning may be lost or changed and so people's goals will not be satisfied. Therefore Grice did not see the maxims just as criteria of meaning but also as rules with normative force. Schiffer extended this normative idea to timeless meaning. Insofar as timeless meaning is

[21] There are of course, exceptions; poetry for example would break some or all of these maxims. However one would *expect* poetry not to conform to these maxims. A poetry reading involves a deliberate flouting of the maxims, with the audience understanding and expecting this.

based on conventions within the linguistic community then it follows that there is a normative reason to obey these conventions. If a person does not obey these conventions then they risk not being understood and so violate the cooperative nature of the conversation.

It should be noted that the rationality of both the conversational maxims and timeless meaning is not that of individual rationality as posited by conventional economics. Its cooperative features are more akin to the notion of "team thinking" put forward by Sugden (1993). In this mode of thinking each person tries to "play their part" in achieving outcomes that are good for all. Actions are not considered in isolation but instead take the actions of others into account. In order to achieve these outcomes there must be a recognition that there is a "team" in existence so that expectations can be coordinated around what is best for this team. In a coordination game for example a person who wishes to coordinate on a Pareto- superior outcome with their opponent will have to take into account what their opponent is thinking in order to successfully coordinate. The same notion of team thinking operates in Grice's notion of the cooperative principle. In this, the team is the speaker and listener who have a mutual interest in forwarding the conversation.

5. Application of Gricean ideas to experimental situations

The application of Grice's ideas to the experimental and decision- theoretic problems outlined in the first two sections have briefly been mentioned. In this section we will examine this subject in more detail. First we will discuss how Grice's theory fits into the experimental setting as a theory of understanding. Secondly we will discuss the use of Grice's conversational maxims in constructing tests of understanding. Finally, there will be a discussion of the differences between Grice's ideas on understanding and the other criteria discussed above.

Grice's theory of meaning provides a criterion for understanding in experiments because experiments tend to involve a conversation between the experimenter on one hand and subjects on the other. It is a one sided conversation in that the experimenter is talking and the subjects try to understand what the experimenter is trying to say about the experiment (although sometimes subjects may ask questions). The "conversation" is not purely oral since much of the

communication takes place using examples, illustrations and practice run- throughs. The content of the communication between the experimenter and the subjects includes the structure of the tasks done in the experiments including (in lottery experiments) the distinguishing of outcomes.

Since subjects are recruited voluntarily to the experiment and, in economic experiments at least, are provided with an incentive, so subjects have an interest in discovering how the experiment works. Also, given that there is no deception in the experiment , the main motivation of the experimenter is to ensure that the subjects do the experiment according to the instructions. This means that the two sides of this "conversation" have a common purpose in making sure that the subjects understand what is going on in the experiment.

Given that the instructions are given in a language which is in the repertoire of the subjects[22], then the "timeless" meaning of the experimenter's utterances should be fairly straightforward. However, there is still a large gap between the timeless meaning of a communication and the meaning which the subjects pick up. This gap in communication is filled by the conversational maxims. The maxims apply because the cooperative principle applies in this case since the subjects and experimenter have a common purpose in the experiment.

It follows that, when designing an experiment, an experimenter needs, at least implicitly, to bear in mind the implications of the four maxims. Any violation of these maxims may result in an unintentionally different meaning being given to the subject. This obviously applies not only to the formal instructions given to subjects but also to examples, tests of understanding and the structure of the experiment. The latter is important since it has been shown (Schwarz 1996) that the layout of the experiment itself can have a significant effect on how people choose.

When discussing the maxims in the context of experiments it must be remembered that Grice's discussion of meaning did not take place in a void but was supported by several givens. This included the context of the conversation (i.e. the experiment), the background knowledge of the participants and common knowledge of relevant items in the experiment. All of these things are important in any discussion of the methods of a particular experiment. As an example, the "context" of the experiment is particularly relevant to discussions about the "realism" of experiments

[22] This is an important point- there has been much informal debate about whether subjects (in the UK) should be used when their first language is not English.

(see Starmer(1999)). However, although important, these factors are not immediately relevant to our discussion here.

The application of the Gricean maxims to economic experiments is already reflected in some of the rules mentioned earlier. As already mentioned, the Quality maxim relates very closely to the general methodological prohibition against deceit in economic experiments. However it also relates closely to the unease which many experimentalists have with the use of problems with hypothetical incentives or with the use of "cover stories". There is quite often the feeling that, by using hypothetical incentives, subjects are not answering "real" problems since the hypothetical incentives can be seen as false information for the subject. Similar objections can be made against the use of cover stories.

The Quantity maxim also has obvious applications. Repetition of phrases in instructions often leads to confusion because it suggests importance where there may be none. By contrast the provision of too little information means that the subject does not understand what is going on and so may not perform correctly in the experiment. However it should be noted that experiments by their very nature "flout" the Quantity maxim in that they do not give all the information necessary to come to a unique solution to the experimental problem. There are always choices which need to be made by the subject[23]. Experimental sessions therefore will always exclude *some* relevant information. However this particular flouting of the maxim should not matter as the subjects *expect* the Quantity maxim to be flouted in this way in an experimental setting.

The third maxim, that of Relation, is also important in economic experiments. All instructions and examples must be relevant to the task in hand and the information must not include anything which may distract the subject. There are also more subtle questions relating to this maxim. The task and instructions must be structured to give a clear direction for the subjects to gain their rewards which requires that the task and instructions must be so designed as to be relevant for this purpose. Relevance is also another argument, together with the Quality maxim, against hypothetical situations or incentives. If these situations are not seen as relevant to the surroundings or context of the experiment then this may impair subjects' understanding of the problem.

[23] Indeed Thaler's (1988) objection to Binmore's (see Binmore 1999) experiment on the Ultimatum game was precisely that by telling subjects to maximise their earnings, Binmore was giving the subjects too much information.

The Relation maxim is also important when discussing whether repetition of tasks, as in the Discovered Preference hypothesis, is a good method for promoting understanding in experiments. While there are some grounds for this in terms of the Quantity maxim (i.e. giving the subject sufficient information) it seems less persuasive when analyzed using the Relation maxim. A subject may be puzzled about the point of continual repetitions of the task since many repetitions in such a short space of time are unusual. Also, as Schwarz (1996) points out, there is a presumption that repetition of identical tasks is redundant so subjects may try to reinterpret the repetitions of the task in a way which may use irrelevant information. It may be difficult to explain these repetitions without giving away the purpose of the experiment.

The final maxim, that of Manner, is likewise of crucial importance in experimental design. The requirements that one should avoid obscurity and ambiguity are obviously important when promoting greater transparency within an experiment. The requirement that one be brief and orderly reflects the desire that experiments should not be too complicated and that instructions should not be too long- winded.

Since Grice's notion of meaning does have relevance to the question of understanding in economic experiments, this means that it is of interest to compare it with the criteria for understanding given above. The four theories chosen , in order, will be those of Binmore and Plott, Broome, Anand and Cubitt et al[24]. The analysis will be in terms of the underlying rationales for the criteria (or lack of criteria!) of understanding and whether the Gricean notion is consistent with them.

The "Discovered Preference" hypothesis has been discussed and critiqued in section 2. From this it is worth noting that the motivating factor behind it is the idea that one can only possibly understand the experiment (allowing for other possible errors) if one is maximising utility. This is based on the idea that a subject's "true" preferences are always consistent with EU and that any observed deviation must be the result of an error, possibly involving understanding. If no other type of error is made, then it is not possible to understand the experiment and not maximise EU.

This contrasts heavily with the Gricean view of understanding. This view does not identify understanding with whether the subject is observed to be maximising expected utility within the experiment. Instead it focuses on the interaction between

[24] We will ignore Gigerenzer's notion as it adds little to the discussion being a variant (from the point of view of understanding) of the Discovered Preference Hypothesis.

the experimenter and the subject. The cooperative principle states that these two agents do try to increase the subject's understanding of the experiment because there is a common purpose, although the motivations are different. However, this process of gaining understanding is independent of the subjects' own process of utility maximization. The method by which a subject gains a prize on having full understanding is not assumed to be under the direct control of the experimenter.

A similar argument can be made against Broome's (implicit) theory of understanding. Given the cooperative principle, if the conversational maxims are fulfilled this should result in the subjects having the same understanding of the problem as the experimenter. Furthermore, the cooperative principle implies that there is no incentive for the subject to deviate from this understanding. A redescription of outcomes by the subject would violate the principle.

For both the Discovered Preference Hypothesis and Broome's ideas it should be emphasised that full understanding by the subject does not mean that the subject would maximise utility in the experiment itself. They may in fact have a completely different idea of the best way to choose (so they may, for example, choose according to Regret Theory). However since *to the subject* the way they choose is the best way, it is perfectly consistent to obey the cooperative principle and violate expected utility. Full understanding and expected utility maximization are two separate and different goals.

At the opposite end of the "understanding" spectrum is Anand's argument. As mentioned above, Anand does not believe that there is any linguistic base upon which we can rationally distinguish between states of the world and so it is reasonable to respecify outcomes in a problem in any way one wishes[25]. Anand bases this belief upon the idea that there is no such thing as a "rational" language which can exclude such redescriptions. However, this is a peculiar position to take. It is not necessary for a language to exclude one set of descriptions and prioritise another set. All that is necessary is that the agent understands, in a particular situation, which is the most appropriate description to use.

There have been a variety of suggestions as to how to do this (see the "natural kinds" literature (Quine 1969, Putnam 1975) for an alternative). However, Grice's analysis provides the most natural analysis for an experimental situation and is a

[25] Although the existence of agreed linguistic conventions may mean that in practice there is an agreed distinguishing of outcomes.

widely accepted approach for defining meaning (and so understanding). If a word has timeless meaning in a language shared by the experimenter and the subject then this implies that this meaning is a convention in the language. If it is a convention then, given the cooperative principle, it is rational for the subject to use that meaning rather than any other meaning. It follows that the correct individuation of states is provided by the experimenter to the subject[26] and that a cooperative subject will follow this individuation. It follows from this that Anand's pessimism is unwarranted in an experimental situation.

The final position is that of Cubitt et al. Their position is essentially pragmatic, based on empirical testing and controls within the experiment. There is no "criterion" as such, just a set of rules of thumb derived either from experience in experimental practice or from contexts outside the experimental or economics arena. The big flaw with this approach is its *ad hoc* structure. There is no particular *theoretical* reason to believe that the bank of tests, instructions and practice rounds given to the average subject will necessarily improve understanding.

Grice's work however fits in very well with this view of understanding. There is an a priori assumption that subjects are trying to understand the experiment (i.e. the cooperative principle), while there are features which allow for controlling the experiment in case the subjects do not understand it (i.e. the conversational maxims). Furthermore there is no link between utility theory in the choices made and understanding; it is possible for a subject to understand an experiment and still fail to maximise expected utility.

This "Gricean" view of understanding therefore can act as a structuring theory for the pragmatism of Cubitt et al. It does not explicitly derive proposals for how understanding can be improved (although the structure may inspire more tests) but it does provide a taxonomy within which experimentalists can work. Furthermore, it provides a normative reason (the cooperative principle) for why it is rational to work on the assumption that subjects do generally try to understand experimental instructions.

[26] It could be said that a "rational" language in this case is simply the one which the experimenter is using!

6. Other applications of Gricean theories to experiments

While there has been no general comparison of Grices ideas to notions of understanding in experimental economics and decision theory, there have been other attempts to discuss the implications of Grice's theories in the general experimental field. Some of these have used Grice's conversational maxims as controls on experiments while others have discussed the implications of violating maxims when looking at some classic "violations" of rationality.

Two works which have explicitly discussed Grice's ideas in an experimental context are those of Schwarz (1996) and Hilton (1995). Their work has been empirical in that they have demonstrated the *empirical* applicability of Grice's maxims to psychological theory and have carried out tests of this applicability. According to Schwarz and Hilton, many experiments which have resulted in what have been acknowledged to be "biases" in the literature are in fact the result of violations of Gricean conversational maxims i.e. in the terms of this paper their understanding differs from that of the experimenter. In general, the idea is that subjects, rather than being irrational, in many cases are actually being rational within a different interpretation of the question. The subject is not being awkward and in fact is obeying the cooperative principle, but the different understanding is the result of the experimenter violating the maxims of conversation.

An example of this is the Base Rate fallacy. This is based on a classic experiment by Kahneman and Tversky (1973) where subjects are each given a description of a person which excludes details of that person's career. This description was said to have been formulated by psychologists and to have come from a group of 100 such descriptions of people of which 30 of the people were engineers and 70 were lawyers (or vice versa depending on the treatment). The subjects were then asked whether that person was an engineer or a lawyer. It was found that there was an overwhelming reliance on the description as opposed to the base- rate for the answer even though there was no explicit clue in the description.

However it has been shown (by Schwarz) that this is not robust to changes in presentation. The assumption of relevance (i.e. the Relation maxim) means that all of the information will be used by the subject even if at first sight it seems irrelevant. The focus on the description and on the role of psychologists in selecting it out are all assumed to be relevant and it is this which causes neglect of the base- rate. However,

when these are controlled then the effect disappears. In this case it is the experimenter who has violated the conversational maxims by focussing on irrelevant information and the subject is trying to make sense of it by assuming that the maxims hold.

Other possible violations of conversational maxims in experiments include leading questions which can lead to "false memory" of issues asked about in the questions. Also important are "assertions of the obvious" where instructions include statements which are obviously true. Rational conversationalists would not make such obvious statements and subjects may start to look for hidden meanings. Repeated questions are also violations of conversational maxims since they seem to invite different responses from those given previously. Many so- called "measurement artefacts" are in fact violations of conversational maxims. For example, the use of open questions may lead to a variety of different answers because of the lack of information. (Schwarz 1996)

These and other biases are to a certain extent an interpretation of the existing evidence in terms of the Gricean Maxims. As both Hilton and Schwarz admit, the maxims are simply another useful tool in the experimental armoury- they should not be seen as a general cause for all experimental problems. Violations of rationality and artefacts in experiments are not all caused by violations of the maxims since the latter tend to have multiple causes. Schwarz and Hilton's general conclusion is that Grice's maxims are a serious methodological issue which needs to be tackled in experiments. For Schwarz and Hilton the main focus is empirical i.e. finding hypotheses for explaining anomalies rather than finding a normative and descriptive theory of understanding as in this paper. However their analysis does give support to the idea that subjects do actually assume Grice's maxims in conversation.

While Schwarz and Hilton's analysis is geared towards empirical hypotheses about experimentation there are other theorists who have used Grice's maxims for more philosophical purposes. The ideas which will be discussed here are those of Todorov (1997) which, if true, would completely overturn the argument made so far in this paper. Todorov's claim is simple- that, not only are Grice's maxims important in experimental studies but also, when violated, they are the cause (contrary to Hilton and Schwarz) of many of the violations of rationality seen in the literature. Todorov's position is more complex than this makes it sound since he subscribes to Cohen's (1981) view that it is not possible to test human rationality as such since

"normativity" is actually derived from human intuition[27]. Human "competence" at rational thought therefore is not in doubt while there may be some errors in human "performance".

However, no critique is offered of the latter views here. Instead we will concentrate on Todorov's claims about Grice's axioms. Todorov makes two claims:

i) It is always possible to find an explanation of "violations of rationality" in terms of violations of Gricean maxims.

ii) Cognitive illusions can be avoided by an "appropriate" representation.

It can be seen immediately that these claims simply replicate the positions which we argued against earlier. Claim (i) could be seen as a variant on the Discovered Preference Hypothesis except that "subject error" is replaced by "violation of the maxims" as the main culprit. Meanwhile (ii) is Todorov's extension of Gigerenzer's opinion which is explained by stating that an "inappropriate" representation is one where the subject is confused and violates one of the conversational maxims. These are both complete turnarounds in the previous arguments and, at first glance, seem to reinforce the Discovered Preference or Broome's/ Anand's position. The claim here is that this conclusion is premature since its reasoning is too extreme.

Todorov's claims are reinforced by two examples; one is the Wason Selection Task (Wason 1968) and the other is the Base Rate Fallacy (Kahneman and Tversky 1973) which we have already discussed. Here we will discuss just the Selection Task as the general arguments used for this also apply to the Base Rate Fallacy and also because there is some limited evidence on Todorov's claims about the Selection Task.

The Selection Task is an experiment which tests subjects' ability to carry out conditional reasoning. The original experiment (Wason 1968) involved a layout of four double- sided cards, each of which had a letter on one side and a number on the other. The layout had the letters and numbers "2", "3", "A" and "B" face up. Given this layout, the subject was told to pick the minimum number of cards to test the

[27] While not presenting a critique of this view it is worth pointing out that, while normative rules do originally derive from human intuition they are *refined* over time. Aristotelean syllogisms, for example, may be derived from intuition but a 100- page proof in modern predicate logic certainly isn't.

conditional statement "If a card has a vowel on one side then it has an even number on the other. The rational choice (according to Wason) was to choose cards A and 3 but most subjects chose the "positive confirming" cards A and 2. The latter card is irrelevant for testing the statement since it cannot falsify it. Since Wason's original paper a whole range of variants of the task have been carried out which have extended this conclusion and have shown some of its limitations.

Todorov makes several claims about the Selection Task but two are relevant here. First of all he notes that, while the Selection Task may be performed badly in many circumstances, there are others such as the "Deontic" tasks (i.e. where the statements tested are normative rules) where subjects perform well (see Cheng & Holyoak 1985). Todorov interprets this as being the result of deontic tasks being more "relevant" than the other versions of the task. Since people are used to following deontic rules it follows that a task involving such rules is more likely to be done well than one which is abstract.

Todorov also makes claims about the interpretation of the rule used. He points out that the English interpretation of the conditional used in the rules is an "if… then" statement. However, in colloquial English this could be misinterpreted as a biconditional or as an existential conjunction, both of which could lead to incorrect results. He claims that it is possible, using a different logical interpretation of the conditional[28], to induce subjects to give the correct answer. This is justified on the grounds of cognitive ease, although he presents no empirical results to justify this.

Todorov's arguments are interesting but not convincing. In an experiment carried out with Bob Sugden (Jones and Sugden 2001) we investigated the Selection Task, attempting to control for precisely these points[29]. The phrasing of the rule was changed to "Every X is Y" which gives the conditional in a more explicit manner. This had the advantage of reinforcing the conditional without eliminating its conditional character. This did not result in the disappearance of the anomaly, showing that the ambiguity of the "If.. then" statement was probably not responsible for the bias. Furthermore, the argument in favour of "Relevance" as what is driving the large amount of correct answers in deontic tasks seems to be false. Experimentally, when we controlled for realism using abstract and "realistic"

[28] Specifically for an item x and features P and Q of x the conditional ($\forall x, Px \rightarrow Qx$) can be interpreted as ($\neg\exists x\ Px\ \&\ \neg Qx$).
[29] Although **not** in reaction to Todorov's claims.

examples with deontic and non- deontic statements in both cases, the deontic statements worked just as well whatever the level of realism. Whatever is driving the success of deontic statements it is not their familiarity[30].

While these results should not be interpreted as being conclusive they do demonstrate that Todorov's arguments are not necessarily correct. However, a more interesting point is that Todorov's theories are in fact testable. Todorov's claims would only hold if it was never possible to satisfy Grice's maxims in the conventional form of the Selection Task- this is far too pessimistic an interpretation and can be shown by empirical testing to be false. Grice's maxims are not an a priori logical block to testing rationality, but rather they are empirical hypotheses which can be tested in the laboratory. It may have been the case that Todorov was correct for the Wason Selection Task. However, this does not detract from the fact that this would have to be tested experimentally rather than assumed a priori and that the presentation and explication of the Selection Task could be changed to eliminate these problems.

Interpreting the evidence here, Todorov's two claims are not universally true. Violations of rationality are not always traceable to violations of Gricean maxims and cognitive illusions cannot always be cured by an appropriate representation. However, aside from the empirical evidence, the second claim may be doubted for other reasons. Why, for example, should one test a subject's ability to carry out conditional reasoning, as in the Selection Task, by giving the subject representationally different but logically equivalent non- conditional statements? It may be the case that the new non- conditional representation allows the subject to answer "correctly" but not all problems are constructed in this way. We are interested in how people use conditionals *in general* rather than just in particular cases and the general failure of reasoning associated with conditionals *when represented as conditionals* is an important fact which cannot be argued away by claiming that a different non-conditional representation should be used.

Rather than completely disposing of the conditional form as suggested by Todorov it may be better simply to rephrase the rule as a more explicit conditional. This fits in with the Gricean maxims (by obeying the maxim of Manner and removing obscurity) and allows a test of conditionality. Doing otherwise leads one to doubt what one is testing- a formulation may be logically equivalent to a conditional but if it

[30] It is also worth pointing out that another of Todorov's claims- that the subjects are really acting like Bayesians as suggested by Oaksford and Chater (1994) has also been proved wrong.

is not presented as a conditional then can one really say that one is testing conditionals?

7. Conclusion

The main thrust of this paper has been to discuss the role of Grice's Conversational maxims within experimental economics. The general conclusion is that they can be used to provide a coherent set of criteria of understanding within experiments. This set of criteria for understanding, it has been argued, makes more sense in the experimental setting than other possible rivals.

Crucially the main argument in this paper has been towards a "commonsensical" view of understanding in experiments. Grice's analysis was precisely that- an analysis of the notion of "meaning" as it is used in ordinary language rather than in a complex philosophical model. One of the arguments he made in favour of the cooperative principle was that this was precisely how people *did* communicate.

This means that the analysis here endorses a pragmatic, commonsense view of understanding as put forward by Cubitt et al. It follows that it is possible to understand an experiment and yet be "irrational". It is also possible for the subject to come to understand the experiment in the same way that the experimenter understands it. Subjects do not repecify the problem in order to fit in with their own (consistent) preferences and there *is* a baseline criterion for understanding by the subject.

The implication of this is that the notion of understanding becomes empirical and practically testable. This contrasts sharply with Plott's (1996) claim that Discovered Preference was a philosophy or means of interpreting the data rather than a testable theory. One of the claims of this paper is precisely that understanding *is* testable and can be tested independently of whether a subject is irrational or not. This also opens up the possibility that Discovered Preference as a theory about learning (i.e. that all subjects will eventually come to understand an experiment through learning and repetition) is also testable. This has been shown through the discussion of repetition and the Relation and Quantity maxims.

While this paper fits in with the Cubitt et al. position in the economics literature it also fits in with the methodological discussion which has been developed by Schwarz and Hilton. While their work has focussed on the problem of how the

Gricean maxims could be used to make sure that experiments are constructed correctly, this paper extends this to looking at ideas about understanding in the experimental economics literature and in particular looking at normative aspects. The Hilton and Schwarz papers could be seen as applications and tests of the "Gricean" ideas about understanding discussed in this paper so the two streams of thought are complementary.

Grice's theory of meaning and our adapted theory of understanding rely on subject and experimenter being in a conversation with each other. It may be argued that this is a special case and that there are many situations where people have to make choices but are not in a conversation. This is a valid point but is should be remembered that Grice's theory is far more general than it appears. A "conversation" (in this context) need not be to specific people, be oral or happen at the same time for the "speaker" and "listener". A diary or a warning on a wall for example, both count as "conversations" in this context. Indeed any communication in society could be brought under the Gricean theory of meaning[31].

It follows that the arguments given in this paper have an importance way beyond that of experimentation. However, the principle aim of this paper is to regularise the notions of understanding in experiments and Grice's theory of meaning provides an excellent means to do this task. Perhaps the best argument in favour of a Gricean view is that it makes most sense in an experimental context compared to the other views. Its view of understanding chimes in with the assumptions that most experimenters make while at the same time allowing for sensible tests of whether subjects are rational or not. Under this view, experiments are a method for testing theories and collecting data rather than vehicles for detecting "appropriate applications" of theories which have *a priori* been declared to be true.

[31] However, natural phenomena such as are investigated by natural sciences do not have their "meaning" defined by Grice.

BIBLIOGRAPHY

Allais M. (1953) “Le Comportement de L’Homme Rationnel devant le Risque : Critique des Postulats et Axiomes de l’Ecole Americaine” Econometrica vol.21, p. 503-546

Anand P. (1993) “Foundations of Rational Choice under Risk” Oxford University Press Oxford

Avramides A. (1989) “Meaning and Mind” MIT Press Cambridge; Massachussetts

Binmore K. (1999) “Why Experiment in Economics?” Economic Journal vol. 109 p. 16 - 24

Broome J. (1991) “Weighing Goods” Oxford; Blackwell

Cheng P. and Holyoak K. (1985) “Pragmatic Reasoning Schemas” Cognitive Psychology vol. 17 p. 391- 416

Cohen L. (1981) “Can Human Irrationality be Experimentally Demonstrated?” Behavioral and Brain Sciences vol. 4 p. 317- 370

Cubitt R., Starmer C. and Sugden R. (2001) “Discovered Preferences and Experimental Evidence of Violations of Expected Utility Theory” Journal of Economic Methodology vol. 8 p. 385- 414

Gigerenzer G. (1996) “On Narrow Norms and Vague Heuristics” Psychological Review vol. 103 p. 592- 596

Grice P. (1989) “Studies in the way of words” Harvard University Press, Cambridge Massachussetts

Hausman D. (1993) “ The Structure of Good” Ethics vol. 103 p. 792- 806

Hilton D. (1995) “The Social Context of Reasoning: Conversational Inference and Rational Judgment” Psychological Bulletin vol. 118 p. 248- 271

Hirsch E. (1988) “Rules for a Good Language” The Journal of Philosophy vol. 85 p. 694- 717

Jones M. and Sugden R. (2001) “Positive Confirmation Bias in the Acquisition of Information” Theory and Decision vol. 50 p. 59- 99

Kahneman D. (1996) “Comment” in K.J. Arrow, G. Colombatto, M. Perlman and C. Schmidt (eds) “The Rational Foundations of Economic Behaviour” Basingstoke; Macmillan

Kahneman D. and Tversky A. (1973) “On the Psychology of Prediction” Psychological Review vol. 80 p. 237- 251

Kahneman D. and Tversky A. (1996) “On the Reality of Cognitive Illusions” Psychological Review vol. 103 p. 582- 591

Lewis D. (1969) “Convention” Harvard University Press Massachussetts

Loewenstein G. (1999) “Experimental Economics from the Vantage- point of Behavioural Economics” vol. 109 p. 25- 34

Loomes G. and Sugden R. (1982) “Regret Theory: An Alternative Theory of Rational Choice under Uncertainty” Economic Journal vol. 92 p. 805- 824

Machina M. (1981) “Rational Decision Making versus Rational Modelling?” Journal of Mathematical Psychology vol. 24 p. 163 - 175

Oaksford M. and Chater N. (1994) “A Rational Analysis of the Selection Task as Optimal Data Selection” Psychological Review vol. 101 p. 608- 631

Plott C. (1996) “Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis” in K.J. Arrow, G. Colombatto, M. Perlman and C. Schmidt (eds) “The Rational Foundations of Economic Behaviour” Basingstoke; Macmillan

Putnam H. (1975) “Mind, Language and Reality” Cambridge; Cambridge University Press.

Quine W. V.O. (1969) “Natural Kinds” in “Ontological Relativity and Other Essays” Columbia University Press

Schiffer S. (1972) “Meaning” Oxford, Clarendon Press

Schwarz N. (1996) “Cognition and Communication: Judgmental Biases, Research Methods and The Logic of Conversation” Lawrence Erlbaum Associates; New Jersey

Sen A. (1985) “Rationality and Uncertainty” Theory and Decision vol. 18 p. 109-127

Smith V. (1982) “Microeconomic Systems as an Experimental Science” American Economic Review vol. 72 p. 923- 55

Starmer C. (1999a) “Experiments in Economics… (Should we trust the Dismal Scientists in White Coats?)” Journal of Economic Methodology vol. 6 p. 1-30

Starmer C. (1999b) “Experimental Economics: Hard Science or Wasteful Tinkering?” Economic Journal vol. 109 p. 5-15

Sugden R. (1993) “Thinking as a Team: Toward an Explanation of Non- Selfish Behaviour” Social Philosophy and Policy vol. 10 p. 69- 89

Temkin L. (1994) “Weighing Goods: Some Questions and Comments” Philosophy and Public Affairs” vol 23 p. 350 - 380

Thaler R. (1988) “Anomalies: The Ultimatum Game” journal of Economic Perspectives vol. 2 p. 195-206

Todorov A. (1997) “Another Look at Reasoning Experiments: Rationality, Normative Models and Conversational Factors” Journal for the Theory of Social Behaviour vol. 27 p. 387- 417

Tversky A. (1975) “A Critique of Expected Utility Theory: Descriptive and Normative Considerations” Erkenntnis vol. 9 p. 163- 173

Tversky A. (1996) “Rational Theory and Constructive Choice” in K.J. Arrow, G. Colombatto, M. Perlman and C. Schmidt (eds) “The Rational Foundations of Economic Behaviour” Basingstoke; Macmillan

Wason P. (1968) “Reasoning about a Rule” Quarterly Journal of Experimental Psychology vol. 20 273- 281