The better toolbox: Experimental Methodology in Economics and Psychology

Daniela Di Cagno
LUISS, Dept. of Economics
Werner Güth
LUISS, Frankfurt School of Finance & Management, Max Plank Society (MPG) Bonn
Giacomo Sillari
LUISS, Dept. of Political Science

Abstract

In experimental economics one can confront a “don’t!”, as in “do not deceive your participants!” as well as a “do!”, as in “incentivize choice making!”. Neither exist in experimental psychology. Furthermore, controversies exist in both fields regarding data collection methods, e.g. play, strategy (vector) method in strategic game experiments, and concerns for external and internal validity, e.g. field versus lab experiments. In addition to touching on these aspects, we suggest ways to enrich the dimensionality of choice data, if possible, when maintaining the revealed-motive approach. Finally, and most importantly, we recommend to elicit not only choice data but also to collect supplementary data shedding light on how participants deliberate before deciding.

JEL: B41, A12, C91, C70
Keywords: methodology, experiments, game theory

1. Introduction

Experimental research in economics and cognitive, social and economic psychology differs, in part because psychology studies human behavior in its entirety whereas economics concentrates on decision making and, in particular, on conscious decision making. By conscious decision making we mean:

- forward looking deliberation, through which one tries to predict the consequences of one's choice options as well as of circumstances beyond one's control (others' choices, random events), and
- with the aim of selecting an option among one's own that has desirable consequences.

This rules out unconscious or purely emotional (so-called system 1) behavior, in which psychologists are nevertheless much interested. Since it is not our topic to discuss why the two fields do not fully overlap, we simply limit our comparison to the experimental methodology concerning their common interest, namely decision making by forward looking deliberation (Kahneman, 2003)

Experimental psychology has a much longer tradition than experimental economics (see Roth, 1995). The awareness of this might have induced experimental economists to proclaim “But we do it differently!” in order to differentiate their discipline from the mere importing of experimental psychology into economics. Whether that was in fact the case (after all it is not bad to learn from other disciplines), or whether a familiar empirical methodology such as experimental research had to be reevaluated, the fact remains that there are remarkable differences between the experimental methodology of economics and psychology. Discussing these differences is not a new topic (see Camerer, 1996, Hertwig and Ortmann, 2001 and 2008, and, for an encompassing philosophical perspective on methodological issues, Guala, 2005) though, like when comparing religions, in doing so one often neglects the huge overlap and emphatically points out minor differences.

We touch on traditional controversial topics (section 2), trying to shed new light in view of new insights. Our main focus is, however, on the issue of how to collect more informative experimental choice data. In section 3, it is illustrated how to enrich the dimensionality of choice data by redesigning (yet preserving their essential characteristics) experimental paradigms like social dilemma or ultimatum games. Section 4 discusses a more fundamental deviation from the revealed motive approach, that, at least in economics, represent the methodological norm. In the revealed motive approach, one observes choice data and tries to infer the motives underlying them. By imposing hopefully reasonable deliberation dynamics on participants and via recording the results of their reasoning steps, choice data can be supplemented by deliberation data, what allows to make more clear-cut inferences about why and how a certain choice has been made. Section 5 concludes.

2. Traditional controversial topics

A lot has been said about traditional controversial topics. So why should they be taken up again? If no such need exists, this might be not so good news for the discipline of experimental research, as it would indicate that too little progress has been made. Let us begin with the issue of *experimenter deception* (see Ortmann and Hertwig, 2002). Psychologists (in particular social psychologists) allow deception, requiring only that, after the experiment is over, participants be debriefed on how they have been deceived. Experimental economists ban deception tout court. The ban of experimental economists rests mostly on the claim that experimenters' reputation that they never deceive their participants is a precious public good.

Progress, here, could mean that this controversy can now be discussed not only in the abstract but also in light of experimental evidence – what, after all, all experimentalists view as decisive. And in fact, Jamison, Karlan and

Schechter (2008) support the hypothesis that participants behave differently after being deceived by experimenters. Such evidence may not suffice to persuade (social) psychologists to sacrifice the convenience of using deception and to think harder on how to collect data without deceiving subjects. But it should at least make them alert that there is a price to be paid for deception, namely lower quality or, possibly, even flawed data.

Of course, experimental economists do not shy away from studying deception (see the recent special issue edited by Gneezy, 2013, and Gneezy, 2005), for instance by implementing deception games in which some of the participants can deceive others. There is no controversy concerning deception games, as the admissibility of such experiments is not questioned by any experimental researcher across the board. This paves the way to a subtler approach to deception, that is the case in which experimenters try and avoid experimenter deception by engaging "subcontractors". What do we mean by that? In addition to normal participants acting in the experimental situation under scrutiny (let us call them “participant participants”) experimenters hire “experimenter participants,” who are incentivized to deceive “participant participants”. Thus, like employers who do not dare to violate labor protection law themselves but allow their subcontractors to do so, experimenters can avoid deceiving their participants themselves via “subcontracting” the deception job to “experimenter participants” (see, for example, Alberti and Güth, 2012, where 51 out of altogether 54 "experimenter participants" deceived their "participant participants", who afterwards revealed no significant deception effect).

Recall that economists see participants' trust that experimenters will not deceive them as a precious public good to be protected. What happens to this "public good argument", then? There are different possible outcomes that may occur whenever a (participant) participant post-experimentally learns how he has been deceived:

(i) after participation in a deception experiment, participants consider new experiments more suspiciously, possibly by discussing their experience with other participants (what might influence future experimental behavior) and irrespective of whether they were deceived by experimenters directly or by "experimenter participants);

(ii) same as above, but only if deceived directly by experimenters and not if deceived instead by an “experimenter participant;”

(iii) no suspicion at all in future experiments.

Let us consider (i) versus (iii). One should concede that just the one aforementioned study (Jamison, Karlan and Schechter 2008) may not provide sufficient evidence for rejecting (iii) and inducing more experimenters (after all, it's not the case that all social psychologists actively use deception) to avoid experimenter deception. In line with Gneezy (2005), we advocate that there should be more experimental tests comparing (i) and (iii). However, in light of the widespread acceptance of (i) in the economics profession, it seems at this point rather unlikely that experimental economists will set up experimental studies of experimenter deception and still try to publish their results in economics journals.

Let us therefore focus on (ii). Claiming that deception by “experimenter participants” will not ruin the public good of participants trusting the instructions but that this occurs only when experimenters themselves deceive presupposes quite sophisticated and subtle reactions of (participant) participants. Nobody denies that being subject to more direct deception can render a participant more suspicious. But would participants be able to clearly distinguish whether deception occurred due to “subcontractors” (“experimenter participants”) on behalf of the experimenters as opposed to it occurring due to the experimenters themselves? After all wasn't it the experimenter themselves who intentionally induced their “experimenter participants” behavior accordingly? In our view, such differences in the mode of deception are too subtle to warrant substantially different reactions. In fact, Alberti and Güth (2012) show in their study that this difference does not surface, although, again, one could argue that more studies and robust findings are needed.

In our view, the fast growing literature of experiments studying “the dark side” of human behavior like dishonesty, spite and envy, sabotage, bribery and corruption etc. (see e.g. Fischbacher and Heusi, 2013), which are usually analyzed without employing experimenter deception, renders the analysis of experimenter deception effects much more subtle and difficult. If participants are annoyed by being deceived and if this has spillover effects for future lab behavior, why should they react so discriminately to deception by experimenters, respectively by other participants? One may argue that Ortmann and Hertwig (2002) have analyzed the evidence of experimenter deception more systematically (than for instance Jamison et al. 2008, or Alberti and Güth 2012 have.) Nevertheless, they did not compare the subtler distinction of participants being deceived by other "subcontracted" participants as opposed to plain experimenter deception with post-experimental debriefing. In view of this much broader perspective, the evidence of experimenter deception seems contested to us.

Let us now briefly touch on other controversies, starting with whether one should (not) use the *strategy method* in sequential decision tasks or even the strategy vector method in strategic games. If elicitation method matters (but so far the evidence is at least partly inconclusive), this would be relevant only when running exploratory “testbed” experiments. Consider the example of testing new institutional designs in the lab before deciding whether to use them in the field. In this case, whether an implementation in the field can be recommended in view of the experimental data collected, could depend on the elicitation method used.

For the experimental testing of “theories”, however, the debate about elicitation methods has no merit. If “theories” do not postulate specific elicitation methods (for instance by ruling out normal form play data in sequential games, or lab data in general) then all possible elicitation methods are applicable to test them. So what should be recommended?

- As far as exploratory “testbed” experiments are concerned, the methodological point about elicitation is valid (but the point remains overstated since it is customarily made not just in connection to such experiments).

- For theory testing experiments, however, the concern should not bother experimental researchers but only theorists. This is because if theory testing should depend on data collection method, then the validity of the theory itself should be conditional on the method of data collection.

A similar argument can be made when discussing *lab versus field experiments*, with Peter Bohm (cf. e.g., Dufwenberg and Harrison, 2008) being one of the pioneers in field experiments. It is not bad scientific practice to specify a theory just for natural field settings or only for highly stylized experimental paradigms and then, if and when confirmed for this limited scope, try and check whether it warrants more ample applicability and generalization. Again, in exploratory experiments this approach can be crucial to fruitful resuslts, whereas for "theoretical experiments" the lab is as good as the field and viceversa, when the theory does not restrict itself to one of the two.

Let us finally consider the issue of *incentivizing participants* (see Ortmann and Hertwig, 2006, Guala 2005) what is standard in experimental economics but not as prevalent in (social) psychology. In our view, most experimenters will agree that providing monetary incentives can

- determine how noisy choice data are,
- crowd in as well as crowd out intrinsic motivation.

The former point is based on the intuition that noise originates from a lack of careful consideration for the consequences of choice, which is less likely to happen the more relevant is the choice under consideration. One could, for instance, argue that monetary incentives will often shift choice making from the emotional and fast system 1 to the slow and deliberate system 2, which would make it a more reasoned choice. But, of course, even substantial monetary incentives may not be sufficient to hinder system 1 reactions (see Grimm and Mengel, 2012). Again, one may ask why the debate is not resolved in the way that experimentalist should propagate, namely by running experiments with “incentivized” and "not incentivized" treatments. This approach would also allow, in "incentivized" treatments, to vary monetary stakes systematically. Since also (social) psychologists try to motivate their participants, for example, via natural field descriptions where such choices are made, the issue of using incentives is related to that of using neutral or naturally framed instructions. Both using neutral and natural instructions present its own specific (dis)advantages, which one could or maybe should try to assess experimentally.

Let us conclude this section by discussing why experimental economists predominantly maintain the tradition of providing an *equilibrium benchmark solution*, based on the assumption of a well-defined Bayesian game, at the same time refraining from attempting to induce such a game for the participants. One of the earliest aversion concepts, later on joined by many or perhaps too many other “aversions”, is aversion to risk. Risk aversion is customarily induced and controlled in experiments by using binary lottery incentives. That is, a participant can win either a low or a high monetary prize, and the number of tokens earned in the experiment monotonically determines the probability of winning the high monetary prize. If the function that links the probability of winning the high prize and the number of tokens earned is made public, the idiosyncratic risk attitude of the participant could then be assumed to be common knowledge. Thus, there exists a method for inducing commonly known risk attitudes and yet this method is rarely used even though one often refers to risk attitudes when accounting for experimental findings (e.g. Kagel, 1995).

Arguing that this method does not achieve what it is supposed to (see Selten et al., 1999) should be seen as discrediting the economic concept of risk attitude as measured by the curvature of the utility of money, rather than the binary lottery method. One either rejects the binary lottery method, e.g. due to the evidence of Selten et al. (1999, 2003), and therefore gives up using expected utility of money when analyzing benchmark solutions, or accepts expected utility of money along with the obligation to use binary lottery incentives for inducing commonly known risk attitudes.

As it is done in Holt and Laury (2002), one could try and make do without properly inducing commonly known risk attitudes by subjecting each participant individually to an additional risky choice task. This will not work, though, and especially so if one fails to inform participants about other participants' answers in those strategic choice tasks for which one requires benchmark solutions. More fundamentally, if experimental economists are accepting that "incentivizing" may fail, should they not accept that “not incentivizing” may work, too? At least they cannot claim that, in their own tradition, they always properly try to incentivize what they implicitly claim to study, as judged by equilibrium benchmarks.

3. The revealed motive approach

Experimental economics has been dominated by the empirical method, as induced by so-called revealed preference theory. In neoclassical economics this means that one assumes well-behaved and consistent preferences, which one tries to infer from observed choices. Even when rationality is being questioned, the approach remains similar. Consider postulating satisficing behavior: one assumes satisficing and then tries to infer aspiration levels from a sufficient number of choice data points. We will refer to this experimental method as the *revealed motive approach*. In the revealed motive approach, one tries to infer motives from choice observations. Of course, scholars often collect only choice data in experimental psychology as well, and with the aim of inferring motives from them. In experimental economics, however, this is done nearly exclusively (Brocas and Carrillo, 2003).

In the next section, we will discuss how the revealed motive approach can be supplemented by other methods, often used in (social) psychology. What we illustrate in this section is a way to enrich the experimenter's toolbox in order to improve the accuracy of the inference of subjects' motives from observed choice behavior. Simply put, the idea is that in many experimental paradigms we may be in a position to elicit more informative choice data without jeopardizing the crucial aspects of the paradigm, e.g. its social dilemma aspect or its benchmark solution.

For dilemma games, the basic idea is to rely on the normal form representation of a hybrid game combining independent (that is, simultaneous) and sequential play, what is obviously related to the strategy method. Unlike Fischbacher and Gätcher (2012), however, we will restrict conditioning on qualitative information. The basic idea can be easily illustrated with respect to a Prisoner's Dilemma game where only qualitative conditioning is possible. In its hybrid version, both players $i = 1, 2$ make a binary choice of $s_i^0 \in \{D_i, C_i\}$ and between adaptations $\Delta_i^=$ and $\Delta_i^{\neq}$ with $\Delta_i^=, \Delta_i^{\neq} \in \{K_i, S_i\}$, with the following intended interpretation:

- $s_i^0 = D_i$: i defects if she does not adapt
- $s_i^0 = C_i$: i cooperates if she does not adapt
- $\Delta_i^=, \Delta_i^{\neq} = K_i$: i keeps s_i^0 after learning that $s_1^0 = s_2^0$, resp. $s_1^0 \neq s_2^0$
- $\Delta_i^=, \Delta_i^{\neq} = S_i$: i switches from s_i^0 to the other alternative after learning that $s_1^0 = s_2^0$, resp. $s_1^0 \neq s_2^0$.

Each player can adapt with probability *p,* hence with probability 1 - 2*p* no player adapts. In this hybrid game the actions that get actually played are:
(i) either no player adapts and hence $(s_1, s_2) = (s_1^0, s_2^0)$ with probability $1-2p$, or (ii), with probability *p*, either player $i = 1$ or $i = 2$ can adapt while the other $j(\neq i)$ cannot, i.e. $s_j = s_j^0$. For player i (who adjusts her initial choice) the final choice s_i is given by

$$s_i = \begin{cases} s_i^0, \text{ if } (s_1^0 = s_2^0 \text{ and } \Delta_i^= = K_i) \text{ or } (s_1^0 \neq s_2^0 \text{ and } \Delta_i^{\neq} = K_i) \\ \qquad s_i \in \{D_i, C_i\}/\{s_i^0\}, \text{ else} \end{cases}$$

i.e., i switches, if
$\left(s_1^0 = s_2^0 \text{ and } \Delta_i^= = S_i\right)$ or $(s_1^0 \neq s_2^0 \text{ and } \Delta_i^{\neq} = S_i)$.

Obviously, an opportunistic player i in a positions to change s_i^0 will always, after one round of repeated elimination of strictly dominated strategies, use $s_i = D_i$ by appropriately tweaking $\Delta_i^=$ and $\Delta_i^{\neq}$. It is less obvious, however, to determine s_i^0, as one could try to “signal” cooperativeness through $s_i^0 = C_i$. But justifying s_i = C_i in this way requires that player j can be induced not to rely on s_j = D_j, and that such triggering of player j is sufficiently likely to be worthwhile for player i.

Let us elaborate in more detail when player i's choice $s_i^0 = C_i$ is not necessarily dominated. Player j may use the appropriate adaptations $\Delta_j^= = K_j$ if $s_j = s_j^0$ or $\Delta_j^{\neq} = S_j$ if $s_j \neq s_j^0$ in order to play $s_j = C_j$. But j will only observe $s_i^0 = C_i$ (and hence be able to condition on it) with probability p, which entails that choosing $s_i^0 = C_i$ rather than $s_i^0 = D_i$ brings about a change from $\left(D_1, D_2\right)$ to $\left(C_1, C_2\right)$ with probability p. This may compensate the loss that i incurs with probability $1-2p$ by using C_i as opposed to D_i in the case in which neither player can adapt. What matters more may therefore depend on the probability p. However, since after the adjustment player j ($\neq i$) will always use s_j = D_j, choosing $s_i^0 = C_i$ would not survive the first round of elimination of dominated strategies. But again, as experimenters one should study this experimentally rather than--like ancient philosophers--by introspection only.

Di Cagno et al. (2014) use the methodology above in a hybrid public good game, with the following implemented parameters: i's payoff is $9 - c_i + .8\left(c_i + c_j\right)$ where for $i = 1,2$

$c_i^0 \in \left\{0,1,...,8,9\right\}$

And

$-c_i^0 \leq \Delta_i^? \leq 9 - c_i^0$

With i, j ≠ 1,2 and i≠j as well as

.

$$? = \begin{cases} = \text{ in case of } c_i^0 = c_j^0 \\ + \text{ in case of } c_i^0 < c_j^0 \\ - \text{ in case of } c_i^0 > c_j^0 \end{cases}$$

This avoids possible demand effects for the three types of adaptations via the same choice sets for all three $\Delta_i^?$ which, of course, depend on c_i^0 and c_j^0. The notation is meant to be suggestive since i may want to increase, respectively decrease his final contribution in case of $c_i^0 < c_j^0$, respectively $c_i^0 > c_j$. The final contributions for either player *i* = 1,2 are

$$c_i = \begin{cases} c_i^0 \text{ with probablity } 1-p \\ c_i^0 + \Delta_i^? \text{ with probability } p, \text{ where } ? = \begin{cases} = \text{ in case of } c_i^0 = c_j^0 \\ + \text{ in case of } c_i^0 < c_j^0 \\ - \text{ in case of } c_i^0 > c_j^0 \end{cases} \end{cases}$$

where at most one player can adapt. Their earnings depend on the final contribution (c_1, c_2) via $9 - c_i + .8(c_i + c_j)$ for $i = 1,2$ with $j \neq i$.

The most utopian hope a player i may entertain is that the smallest positive contribution $c_i^0 = 1$ trigger (with probability *p*) the change of player $j(\neq i)$ most rewarding for *i*, i.e. via $\Delta_j^+ = 9$ from $c_j^0 = 0$ to $c_j = 9$. Thus, $c_i^0 = 0$ dominates without the need to iteratively delete dominated strategies, when $9 \geq 9p + (1-2p)\left(8 + \frac{4}{5}\right) + p\left(8 + \frac{4}{5}10\right)$ or $p \leq 1/37$. This is true since i earns 9 for sure via $c_i^0 = 0$ resulting in $c_i = 0 = c_j$ whereas via $c_i^0 = 1$ he earns

- 9 with probability *p,* when he can adjust and
- $8 + \frac{4}{5}$ with probability 1-2*p*, when neither can adjust but triggers c_i = 1 and c_j = 0
- $8 + \frac{4}{5}$ 10 with probability *p* when the other can adjust and does so fulfilling i's most utopian hope, i.e. c_j = *9*.

This illustrates that with a sufficiently small probability parameter p even the most utopian beliefs about the other's adaptation do not question universal defection in the sense of $(s_i^0 = 0 \text{ and minimal } \Delta_i^=, \Delta_i^+, \Delta_i^-)$ being dominant irrespective of the behavior by the other player $j(\neq i)$.

Thus, when wanting to preserve the immediate dominance of universal defection, one has to implement experimentally a rather small p (in the situation above p = 2.5%) or accept the more demanding concept of once repeated elimination of dominated strategies.

In a methodologically related study, Felli et al. (2014) illustrate a similar possibility for the ultimatum experiment by allowing for concession making. More specifically, both players, proposer and responder, choose a vector of non-decreasing offers (respectively non-increasing acceptance thresholds rather than only one offer, respectively acceptance threshold where game theoretically the final demand, respectively acceptance threshold is decisive[1].) Nevertheless the multi-dimensional bargaining choices might help to distinguish what partners are actually hoping for and what they are – more or less grudgingly -- would accept finally.

One way to discuss the revealed motive approach is to consider the space in which predictions are made. The revealed motive approach tells us whether observed behavior is or is not in line with given motives: if a motive is confirmed, then prediction works well in action space. Even when prediction does not work well in action space, it may be the case that the a certain motive is satisficing to a degree that it hardly reacts to deviations from, say, optimal behavior in action space. In such cases poor predictions in action space may be neglected since in motive or payoff space such deviations are very minor, i.e. one relies on the nearly optimal choices or nearly equilibrium payoffs (see Harrison et al., 2005 and the literature inspired thereof). In

[1] Since Felli et al. (2014) in some treatment allow the next level of demand and acceptance threshold to be used only when at last one party conceded, earlier than the last demand, respect acceptance threshold have only some influence in these treatments.

particular, highly significant effects in action space may only cause very minor effects in payoff space, fot instance because of rather flat payoff curves.

Imagine to be very successful at “predicting optimality” in both action and payoff space by assuming not only well-behaved preferences but also appropriate beliefs concerning the circumstances which are beyond one's control. Does this actually explain anything? When we impose exogenous preferences and beliefs whose rational implications are in line with observed choice and belief data, we are essentially merely transforming the question "why such outcomes?" into the question "why such preferences and beliefs?" What is even more important is that concepts like "well-behaved preferences" and "well-structured beliefs" are not in line with the cognitive psychology of *homo sapiens* although they make perfect sense for the non-existent (except in academic papers) species of *homo oeconomicus*. This discrepancy looms even larger if we realize, as we have pointed out above, that the exogenous, artifical position of preferences and beliefs allows for good prediction but avoids offering good explanation. Let us be clear about it: the concept of commonly known rationality is an informative and important philosophical exercise and rational choice reconstructions of a given behavior can be very enlightening. Nevertheless, neither notion accounts for the way in which we make up our minds and determine our choices, nor do they provide a basis for improving human decision making. Actually, in cognitive space, rational choice explanations are bound to fall far off the mark, since any component of decision making that is problematic for actual decision makers is either exogenously given or it is neglected. One therefore refers to them by as-if explanations.

It is, of course, easy to criticize but much more difficult to provide an alternative approach. As rational choice offers only a terminlogy and methodology (one can only generate choice and payoff functions when preferences and beliefs are given), what we will offer in its stead is no readily available choice algorithm. Here the necessary input to predict choice behavior has to be generated by the decision maker and is not simply supposed to be exogenously given. Such behavior can be still reasonable but it is at best boundedly rational. Like rationality, it is consequentialist: one chooses among options by anticipating their consequences. But when engaging in forward looking deliberation, one is cognitively constrained. Rather than optimizing, e.g. maximizing payoff, one tries to find some satisficing choice by engaging in a dynamic deliberation process.

4. Deliberation and choice data

For the rational choice approach one needs well-behaved preferences and (Bayesian) beliefs which one considers as exogenously given. This requires Bayesian reasoning – one has to assign numerical probabilities for each and every event and to process probabilities according to probability theory – and intra-personal payoff aggregation. The latter means that one assumes a unitary actor who

- considers the consequences of his choice behavior for each and every constellation of circumstances beyond his control, e.g. the behavior of co-players and random events,
- weighs his evaluation (based on her well-defined preferences) of these consequences by the probability of each such constellation and
- optimizes the sum of these probability weighted evaluations.

Since economists are usually reluctant to engage in inter-personal payoff comparisons, it is surprising that they so readily accept intra-personal payoff comparisons in the form of adding up the probability weighted evaluations by different alter egos of the same individual. It seems that if the former is considered difficult, the latter should be causing similar difficulties. But as illustrated by the popularity of prospect theory, psychology also readily accepts intra-personal payoff aggregation.

To align rational choice predictions with observed behavior, one-step-away deviations from rational choice have been allowed, by and large in the biases and anomalies program (see Tversky and Kahneman, 1974), e.g. via weakening or substituting one of the axioms on which rational choice is based. Later on, progress has been made by including additional motives like other-regarding concerns, ethical inclinations etc. which are partly captured by so-called aversions trading off the achievement levels of different motives (see Güth, 2008). Since this has been frequently done in both economics and psychology, we are left wondering about the success and wide acceptance of such attempts. Of course, when one bias, anomaly, aversion, etc. can account for many important stylized findings, this is very informative. Still, it only transforms the question "why such behavior?" into the question "why such bias, anomaly, aversion,…?" One might try to give evolutionary reasons, what renders such explanations much more demanding. In particular, it does not suffice to mention a possible situation in our evolutionary history (see partly the contributions in Barkow, Cosmides and Toby eds., 1992), where this bias, anomaly, aversion,… could have been fitness enhancing. One should produce more convincing arguments, for example prove more rigorously evolutionary stability and robustness of such behavioral traits.

Let us, however, focus on a more fundamental aspect. Whereas in (experimental) economics the rational choice approach, based on exogenously given preferences, is still dominating, in spite of prospect theory and related concepts, psychologists predominantly see forward looking decision making as a deliberation process. We agree, and suggest a process model with the following steps:

- first, one has to cognitively perceive one's decision task (mental modeling); this mainly requires causal links through which one can predict how one's own choice and uncontrollable circumstances, e.g. the choices by others and random events, affect the achievement levels of one's action goals,
- rather than considering the Bayesian universe of all uncontrollable circumstances, one focuses on a few relevant ones, called scenarios, which one does not dare to neglect without necessarily being able to specify numerical probabilities for them,

- forms aspiration levels for each action goal and each self-generated scenario which
- one tries to satisfice, typically by considering one choice option after the other where search will stop in light of one's satisficing success[2].

There may be other frameworks to accommodate deliberation dynamics. What we want to illustrate by such a process structure is that experimentalists are not restricted to choice data but can impose such deliberation dynamics, for instance by using a suitable software. By forcing participants to use the software, one can observe the results of all reasoning steps and feedback loops and hopefully collect data which, together with choice data, help infer more clearly how and why a certain choice has been selected. Of course, scenario generation, aspiration formation and satisficing search data are more reliable when not only choice behavior but also these cognitive steps are appropriately incentivized (see, for example, the comparison of incentive and no-incentive treatments by Güth et al., 2009).

More difficult to obtain are direct data of mental modelling. Using the mouselab technique to record which parameters, characterizing a choice task, are actually when and in which order retrieved might provide some hints what finally influences choice behavior. Similarly, payoff calculators and other software devices can indicate which causal relationships have been considered. Eye-tracking, controlling decision times and physiological reactions and brain scanning during information retrieval can provide evidence how one reasons about choice making. Direct insights in reasoning processes, at least as far as conscious aspects are concerned, can be obtained by asking individual participants to reason loudly what is recorded and can be analyzed, video- and audio-recording discussions of groups of individuals who, as a group, have to jointly reason and decide what to do.

[2] After having searched long before finding a satisficing action one will likely stop search whereas finding a satisficing action immediately might induce aspiration adaptation (e.g. Sauermann and Selten, 1962), scenario updating, e.g. in the form of adding new scenarios.

Of course, the usual arguments apply that this might change the process structure of forward looking decision deliberation. But to obtain deliberation data in addition to choice data will always require a price. Actually, it seems that journals are more open to fancy methods like brain scanning and medical priming and measurements what seems to discourage attempts to use loud reasoning or to deliver the outcome of reasoning steps by imposing some reasoning dynamics, as indicated above. Of course, the latter is cumbersome, even when not any longer requiring transcripts and content analysis but only independent assessments of causal influences. Altogether all the methods might help to finally provide qualitative and quantitative evidence to more clearly understand what influences the ways how we cognitively perceive more or less demanding decision tasks and generate our choices in the light of their anticipated consequences.

That such data can more clearly discredit certain hypotheses, for example, that sequential games are solved by backward induction, can be easily illustrated, for instance, by an ultimatum experiment with the proposer not knowing the conflict payoff of the responder. When the proposer chooses an offer without retrieving this information, e.g. even though information retrieval is costless, then sequential rationality requiring backward induction is definitely rejected (Berninghaus et al., 2012 provide related evidence for stochastic alternating offer games).

Our suggestion to apply multi-dimensional decision formats or to gather deliberation data in addition to unidimensional choice data wherever suitable, i.e. mainly for theory, respectively hypothesis testing experiments, should be clearly distinguished from experimental methods which elicit correlates of decision deliberation in addition to choices like brain scanning, eye tracking, physiological measures, e.g. testosterone (dynamics), and decision times. As decision times are often always recorded, e.g. when using z-tree (Fischbacher, 2007) there is the risk (similar to reporting gender effects only when they exist) to report them in a very biased way. In our view, one should therefore rely on response time, measuring how long it takes to make and confirm a choice when this is what one wants to study experimentally (see for instance Spiliopoulos and Ortmann, 2014). Clearly, such correlates can be quite informative on how relevant choices are and on which aspects of the decision tasks are to be deemed as crucial. Nevertheless, there exists by and large no direct connection between such indicators and how and why one reasons and decides. Whereas multi-dimensional choice formats may reveal such connections more clearly what still has to be demonstrated more systematically, we are quite confident that deliberation- and choice-data come at a price (the to some extent or other imposed decision dynamics) but renders conclusions how and why certain choices are made much more reliable.

5. Conclusions

Experimental research in economics and psychology has a lot in common

- by analyzing the same or similar paradigms like social dilemma games, payoff allocation games, e.g. the early reward allocation games in social psychology, based on dictatorial allocation, and other experimental workhorses,
- by having to cope with the same or similar difficulties, for example, how entitlement (experimental roles are earned rather than randomly assigned) and stakes (lab rewards do not always correspond to what is at stake in the field) can question what is experimentally observed,
- by sharing software, data, participants and even labs where in the latter case the experimenters from economics still might want to rule out (experimenter) deception.

Altogether the longer tradition of experimental research in psychology has been enormously inspired and enriched by the impressive boom of experimental economics. On the other hand psychologists may already have understood the limits of experiments as a method of empirical research what still has to be realized by experimental economists. Running field experiments may often help but cannot avoid the limitations of experimental research altogether. Statistical, questionnaire, panel data etc. are often not as specific as one's research questions require but can be of much superior quality, e.g. in view of external validity.

What we argue for is to intensify the mutual exchange and inspiration by

- reducing the claim of different methodology to its core elements, for example, regarding incentivization and deception of (participant) participants,
- enriching the dimensionality of choice behavior for the sake of more clear-cut inferences from choice observations,
- collect also deliberation data in addition to choice data.

It has been demonstrated that this can be done. But only when many do so can one hope that improved experimental methodology will crucially enhance our understanding of how and why participants have chosen what we observe. As Gneezy (2005) stressed out: ”people are not indifferent to the process leading up to the outcome”.

References

Alberti, F. and Güth, W. (2012), “Studying deception without deceiving participants: An experiment of deception experiments”, *Journal of Economic Behavior and Organization*, 93, 196-204.

Barkow, J., Cosmides, L. and Toby, J. (eds.) (1992), *The adapted mind:Evolutionary Psycology and the generation of culture*, New York, Oxford University Press.

Berninghaus, S.K., Güth W., Schosser,S. (2014), “Backward induction or forward reasoning? An experiment of stochastic alternate offer bargaining”, *International Game Theory Review,* 16 (1).

Brocas, I. and Carrillo, J. D. (eds.) (2003), *Economics and Psycology*, New York, Oxford University Press.

Brandts, J. and Charness, G. (2011), “The strategy versus the direct-response method: A first survey of experimental comparisons”, Experimental Economics, 14 (3), 375-398.

Camerer, C. (1996), “Rules for Experimenting in Psycology and Economics, and Why They Differ”, in Understanding Strategic Interaction: Essays in Honor of Reinhard Selten, Springer, New York, 313-327.

Di Cagno, D., Galliera, A., Güth, W. and Panaccione, L. (2014), ”A Hybrid Public Good Experiment Eliciting Multi-Dimensional Choice Data”, CESIEG WP. 1.

Dufwenberg, M. and Harrison, G.W. (2008), “Peter Bohm: Father of field experiments”, *Experimental Economics*, 11, 213-220.

Felli, C., Güth, W., Mata-Perez E., Ponti G. (2014) “ Asymmetric Concession Making: An Experimental Study“, mimeo.

Fischbacher, U. (2007) "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics* 10(2):171-178.

Fischbacher, U. and Gaechter, S. and Fehr, E. (2001) "Are people conditionally cooperative? Evidence from a public goods experiment", *Economics Letters*, 71(3):397-404.

Fischbacher, U. and Gaechter, S. (2010), “Social Preferences, Beliefs and the Dynamics of Free Riding in Public Goods Experiments”, *American Economic Review,* 100 (1), 541-556.

Fischbacher, U. and Gaechter, S. (2012), “The behavioral validity of the strategy method in public good experiments, *Journal of Economic Psycology*, 33(4), 897-913.

Fischbacher, U., and Föllmi-Heusi, F. (2013) "Lies in disguise—an experimental study on cheating." *Journal of the European Economic Association* 11(3), 525-547.

Gneezy, U. (2005), “Deception: The role of consequences”, *American Economic Review*, 95 (1), 384-394.

Gneezy, U. (ed.), (2013), “Special issue: Deception, Incentives and Behavior”, *Journal of Economic Behavior and Organization*, 93,196-413.

Grimm, V. and Mengel, F. (2012), “Let me sleep on it: Delay reduces rejection rates in ultimatum games”, *Economic Letters*, 111 (2),113-115

Guala, F. (2005), *The Methodology of Experimental Economics,* New York: Cambridge University Press.

Güth, W., Levati M.V., Ploner M. (2009), An experimental analysis of satisficing in saving decisions, *Journal of Mathematical Psychology*, 53, 265-272.

Güth, W. (2008), "(Non) Behavioral Economics- A Programmatic Assessment", *Journal of Psycology*, 216 (4), 245-254.

Güth, W., and Kliemt, H. (1994) "Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes." *Metroeconomica* 45 (2): 155–87.

Jamison, J., Karlan, D. and Schechter, L. (2008), "To Deceive or not to Deceive: The Effect of Deception on Behavior in Future Laboratory Experiments", *Journal of Economic Behavior and Organization*, 68 (3-4), 477-488.

Harrison, G.W., Johnson, E., McInnes, M.M. and Rutstrom, E. (2005), "Risk Aversion and Incentive Effects: Comment", *The American Economic Review*, 95 (3), 897-901.

Hertwig, R. and Ortmann, A. (2008), "Deception in Social Psycological Experiments: Two Misconceptions and a Research Agenda", *Social Psycology Quarterly*, 71,3,222-227.

Holt C.A. and Laury, S. K. (2002), "Risk Aversion and Incentive Effects", *American Economic Review*, 92,1644-1655.

Jamison, J., Karlan, D. and Schechter, L. (2008), "To deceive or not to deceive:The effect of deception on behavior in future laboratory experiments", *Journal of Economic Behavior and Organization*, 68,477-488.

Kagel, J. H., "Auctions: A Survey of Experimental Research," in J. H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics,* Princeton: Princeton University Press.

Kahneman, D. (2003), "Maps of Bounded Rationality: Psychology for Behavioral Economics", *The American Economic Review*, 93(5), 1449-1475.

Ortmann, A. and Hertwig, R. (2001) "Experimental practices in economics: A methodological challenge for psychologists?"*Behavioral and Brain Sciences*, 24 (3)383- 403

Ortmann, A. and Hertwig, R. (2002), "The cost of Deception: Evidence from Psycology", *Experimental Economics*, 5, 111-131.

Ortmann, A. and Hertwig, R. (2006), "Monetary incentives: Usually Neither Necessary nor Sufficient", Cerge-EI wp.307,1-17.

Ortmann, A. (2010), "The Way in which an Experiment is Conducted is Unbelivably Important": On the Experimentation Practices of Economists and

Psycologists, Discussion Paper 2010/06, The University of New South Wales Australian School of Business.

Roth A. (1995), “Bargainig Experiments”, in *Handbook of Experimental Economics*, J. Kagel and A. Roth, eds., Princeton, Nj, Princeton University Press.

Sauermann ….and Selten, R. (1962),” Auspruchsanpassungtheorie der Untmehmung”, *Zeitschrift fur die gesamte Staatswissenschaft*, 118, 557-597.

Selten R., Sadrieh, A. and Abbink, K. (1999), “Money does not induce risk neutral behavior, but binary lotteries do even worse”, *Theory and Decision*, 46, 211-249.

Selten R., Abbink, K., Buchta, J. and Sadrieh, A., (2003), “How to play 3x3 games. A strategy method experiment”, *Games and Economic Behavior*, 45, 19-37.

Spiliopoulos, L. and Ortmann, A. (2014). "The BCD of Response Time Analysis in Experimental Economics" *Available at SSRN 2401325*.

Tversky, A. and Kahneman, D. (1974), “Judgement under uncertainty: Heuristics and biases”, *Science*, 185, 1124-1131.