

“Methodological discussion, like spinach and calisthenics,
is good for us” (Paul Samuelson (1963, p. 231)

spinach : Pronunciation: 'spi-nich Function: noun

Etymology: Middle English spynache, Middle French espinache, ...

1 : an Asian herb (*Spinacia oleracea*) of the goosefoot family

cultivated for its edible leaves which form in a dense basal rosette

2 a : something unwanted, pretentious, or spurious b : an untidy overgrowth

- spin·ach·like /-nich-"Ilk/ adjective

- spin·achy /-ni-chE/ adjective

calisthenics: Pronunciation: -niks Function: noun plural but singular or plural in construction

Etymology: Greek kalos beautiful + sthenos strength

1 : systematic rhythmic bodily exercises performed usually without apparatus

2 usually singular in construction : the art or practice of calisthenics

Completely unrelated: www.journalsfeedback.org

Money, Lies, and Publications

Andreas Ortmann

*Center for Economic Research and Graduate Education,
Charles University
Economics Institute,
Academy of Sciences of the Czech Republic
Prague, Czech Republic*

References:

- Hertwig & Ortmann (2001), Behavioral and Brain Sciences
[target article and commentaries and response]
- Hertwig & Ortmann (2002), in Carrillo & Brocas (eds) OUP
- Ortmann & Hertwig (1996, 1997), American Psychologist
- Ortmann & Hertwig (2002), Experimental Economics
- Ortmann, Hertwig & Harrison (in prep)

Also:

- Camerer & Hogarth (1999), Journal of Risk and Uncertainty
- Gigerenzer, Harrison, Starmer - various articles on methodology

A surprising fact?

- Two related disciplines, economics and corresponding areas in psychology (in particular behavioral decision making), have very different conceptions of what constitutes good experimentation (e.g., use of financial incentives, use of deception, use of repeated trials, use of [abstract] scripts, control of subject characteristics such as gender, social status, cognitive abilities, etc.)
- Why?
- *Disclaimer:* This talk is only about research in social psych, bdm, and research on higher mental processes.

A little history lesson ...

- At least three diverging models of investigative practice in early modern psychology ...
- The Wundt style of experimentation ...
 - Experimenter and participant knew each other, often as long-term collaborators and friends
 - The role of psychological data source was considered to require more psychological sophistication than the role of experimenter
- => Experimentation was a repeated game, involved trust, motivation, expertise. Didn't require financial incentives, deception, etc.
- ... and how it came undone (Danziger 1990)
 - Dramatic shift in experimental practices from the 1920s to the 1940s in the USA
 - Wundt style of experimentation replaced by treatment group experiments, probably to legitimize psychology's practical utility to educational bureaucrats

A little history lesson ... (2)

- The consequence: Experiments as multilevel “game” played intergroup and intragroup (Ortmann & Hertwig 2002)
- The intergroup relationship “has some of the characteristics of a superior-subordinate one Perhaps the only other such one-sided relationships are those of parent and child, physician and patient, or drill sergeant and trainee.” (Schultz 1969, p. 221).

- “Most [subjects], as a matter of fact, are carrying on a train of psychological activity ... ‘Where did I see that man before? -- What is he getting at anyhow? How stupid that experimenter looks! -- What a loud necktie! How stupid must he think I am?’”
Rosenzweig (1933)
- “The fact that the experimenter controls the information available to the subject and that he never reveals completely what he is trying to discover and how he will judge what he observes -- this feature gives the experiment much of its character as a game or contest. It leads to a set of inferential and interpretive activities on the part of the subject in an effort to penetrate the experimenter’s inscrutability.” Riecken (1962)

A little history lesson ... (3)

- Psychologists - standards are very much laissez-faire, allowing for a wide range of acceptable practices
- Economists - standards are rigorously enforced
- Why?
 - “the new kid on a hostile block” (Lopes 1994; Roth 1995)
 - concentrated all the way through 50's, 60's, 70's and 80's in about half a dozen places (e.g., Arizona, Caltech, College Station, Pittsburgh, Frankfurt, ...), facilitating the codification of methodological practices (e.g., Smith 1976, 1982)
 - guided by a (more or less) unifying theoretical framework (less variability in research goals?)

Do these differences matter?

- One cannot talk about experimental results without talking about design and implementation of experiments! In other words, “all tests of theory require auxiliary hypotheses to implement them. Consequently, all falsifying outcomes can be dismissed by arguing that it is the auxiliary hypotheses, not the theory that must be rejected.” (Smith 2001)
- “Without being too personal about it, you probably know that xxx is a fanatic who, many feel, is likely to try to design after design until he gets an effect to go away.” (Anonymous referee of our BBS target article about an experimentalist whose work we discussed prominently and deservedly.)

Do these differences matter? (2)

- “Control is the hallmark of good experimental practice, whether it be undertaken by economists or psychologists.”
(Harrison & Rutstroem 2001)
- Economists reduce uncertainty by defining the demand characteristic of the situation as that of performance (script, financial incentives), allowing participants to gain experience with the situation (repetition), and reducing second guessing (no deception). Psychologists leave room for uncertainty by not clearly defining the demand characteristic of the situation (no script, no payment), interpreting people’s one-off response as indicative of their general competence (no repetition), and inviting second-guessing.

Do these differences matter? (3)

- (Hypothesis 1) A key to understanding the variability of empirical results in psychology (error variance) is the relative laissez-faire attitude of psychologists toward experimental control in their implementation of experiments as regards the use of financial incentives and the use of deception.
- (Hypothesis 2) A key to understanding the variability of empirical results in economics is the relative laissez-faire attitude of economists toward experimental control in their implementation of experiments in areas such as problem representation, control of subject characteristics, “field referents”, etc.
- Let’s look at financial incentives and deception ...
 - (fast, inexpensive, convenient)

What do we mean by financial incentives?

- Performance-based participant payments (not necessarily monetary)
[Note that performance-based participants payments require the existence of a performance standard.]

Use of financial incentives in economics

- Economists almost always employ financial incentives
- In all experimental studies published in AER 1970 - 1997 subjects were paid according to performance (Camerer & Hogarth 1999)

[notes]

Analysis included only studies

- in which some kind of performance criterion was used
- studies where participants were provided with explicit choice scenario involving monetary consequences

Analysis did not include studies

- where experimenters instructed participants that there were no right or wrong answers
- where we couldn't identify clear-cut performance criteria

Use of financial incentives in psychology (especially in areas such as jdm)

- Psychologists rarely use financial incentives
- In only 26 % of 186 experimental studies published in JBDM 1988 - 1997 were subjects paid according to performance (Hertwig & Ortmann 2001)
- Analysis identified upper bound of use of financial incentives in psychology
 - JBDM publishes articles by psychologists, economists, management sciences
 - Experimental economists on editorial board and frequent contributors
 - Subtract those studies where an experimental economist involved, down to 22 %
 - Subtract those studies by (few) psychologists who routinely use financial incentives, down to 15%
- Financial incentives were used in only 3 out of 106 Bayesian reasoning studies in various social psych, cog psych, and jdm journals (Hertwig & Ortmann 2001)

Use of financial incentives in psychology (especially in areas such as jdm) (2)

- Financial incentives are not the norm in behavioral decision making
- Should they be? Do financial incentives matter?
 - “Our subjects are the usual middle class achievement-oriented people who wish to provide [maximal performance]”
(Dawes 1996, p. 20)
 - “Psychologists presume subjects are cooperative and intrinsically motivated to perform well.” (Camerer 1995, p. 599)

Do financial incentives matter?

- Psychologists, for the most part, believe/claim this is not so. “Experimental findings provide little support” for the view that “observed failures of rational models are attributable to the cost of thinking and thus will be eliminated by proper incentives”? (Tversky & Kahneman 1987, p. 90)
- Economists are convinced this is so:
 - Smith & Walker (1993) survey of 31 studies
 - Financial incentives, while not guaranteeing optimal decisions, in many cases bring decisions closer to the predictions of the normative model.
 - Financial incentives reduce data variability
 - Opportunistic sample

Do financial incentives matter? (2)

- Camerer & Hogarth (1999) analysis of 74 studies of “judgements and decisions”, “games and markets,” and “individual choice”
 - Across all three research domains,
 - in 45 % of studies financial incentives did not make a difference
 - in 40 % percent they had positive effect
 - in 15 % effects were negative
 - Conditioned on research domains different picture emerges, e.g., for jdm
 - in 18 % of studies financial incentives did not make a difference
 - in 53 % of studies they had positive effect
 - in 29 % effects were negative

[see Figure 1 of Hertwig & Ortmann 2002]
- Opportunistic sample

Do financial incentives matter? (3)

- Relevant subset of 10 studies in JBDM sample (that either compared payment or nonpayment condition or different payment schemes) showed:
 - in clear majority of cases where financial incentives made difference, they improved participants' performance (and, in fact, did so in exactly the way Smith & Walker (1993) claim they do: by bringing decisions closer to the predictions of normative models and by reducing error variance substantially)
 - in two cases (of which one was compromised by methodological problems) they led to impaired performance
 - in a few cases they did not affect performance
- Non-opportunistic sample

Do financial incentives matter? (4)

- Harrison & Rutstroem (forthcoming) study of 40 contingent valuation studies
 - Strong evidence for “hypothetical bias” in value elicitation methods
- Harrison (1994) re-examined some of Kahneman & Tversky’s studies on cognitive illusions that used financial incentives find that
 - most of these studies lack payoff dominance
 - in re-designed experiments observed choice behavior is in line with the predictions of economic theory
- Recent studies by Parco et al. (forthcoming), Holt & Laury (forthcoming), ...

Do financial incentives matter? (5)

- There seems to be agreement on at least the following points:
 - Financial incentives matter more in some areas than in others
 - In jdm, they seem to matter more often than not
 - The effects, in jdm and elsewhere, seem to be three-fold
 - financial incentives move the data closer to the
game theoretic (decision theoretic) prediction
 - financial incentives reduce the variability of the data
 - financial incentives have dramatic impact on risk aversion

Do financial incentives matter? (6)

- We know little about the substitutability of “capital” and financial incentives:
 - Smith & Walker (1993) argue that increases in participant experience often improves performance at rates that are equivalent to large increases in monetary rewards
 - Camerer & Hogarth (1999) make a similar but more general argument. Eckel (1999) provides an interesting empirical example.

Policy recommendation

- Submit experimental practices to experimentation
 - When evidence suggest that financial incentives (don't) matter, (don't) use them
 - When evidence is mixed (or not existent), accord financial incentives the status of a treatment variable ["do-it-both-ways rule"]
- In other words, experimental practices ought to be empirically grounded. (Note that this policy recommendation, while implicitly critical of experimental practices in psychology, also questions current practices in experimental economics.)
- Not just applicable to financial incentives! [also repetition, etc.]

What is deception?

- Agreement that intentional misrepresentation of the purpose of an experiment or any aspect of it constitutes deception: *“There is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not.”* (Hey, 1998)
- Agreement that it is not feasible, and desirable, to acquaint participants in advance with all aspects the research being conducted (e.g., research hypotheses): *“Absence of full disclosure does not constitute intentional deception.”* (Baumrind, 1985)
- What about violations of default assumptions?

Use of deception in economics?

- Taboo. Always was. Still is.
(There's probably not more than half a dozen published articles that employed it.)
- Why?
 - “The researcher should ... be careful to avoid deceiving participants. Most economists are very concerned about developing and maintaining a reputation among the student population for honesty in order to ensure that subject actions are motivated by the induced monetary payoffs rather than by psychological reactions to suspected manipulation.” (Davis & Holt, 1993)
 - “From a long-range point of view, there is obviously something self-defeating about the use of deception ... Our potential subjects become increasingly distrustful of us, and our future relations with them are likely to be undermined.” (Kelman, 1967)

Use of deception in psychology?

- Psychologists often use deception (as do people in related fields such as marketing).
- [Figure: Use of Deception]
- Deception clearly not a last-resort strategy!
- Compare to APA Ethics Code, 1992:
“Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study’s prospective scientific, educational, or applied value and that equally effective alternative procedures that do not use deception are not feasible.”

What are the effects of deception?

- Attitudes and feelings
 - Suspicion
 - Experimental performance
 - *Direct vs. indirect effects*
-
- Our method: Search for research articles (PsyINFO)
 - Key words: deception in combination with suspicion, expectation, prebriefing, forewarning, experimental history, etc.

What ...? (2) Attitudes and feelings

Two opposing views

- Participants reported that deception experiments are more enjoyable and beneficial than those not using deception.
- vs.
- Participants with a history of participation in deception studies reported caring less about understanding and following instructions. Deception is, next to safety concern, the most frequently mentioned reason for withdrawing from an experiment.

What ...? (3) Suspicion

- Stang (1976) analyzed studies that reported the percentage of suspicious participants.
- About one third of the studies reported that more than 50% of the participants were suspicious.
- For the reported period between 1954 and 1973, the percentage of suspicious participants increased as a function of time, $r = .76$. This correlation tracks the increase of the use of deception in social psychology during that time.

What ...? (4) Experimental performance

- Post-hoc identification of suspicion:
Compared to unsuspecting participants, suspicious participants conformed less in over two-thirds of the studies (medium to large effect size)
- Experimental history and deception:
 - Firsthand experience with deception affects performance, while mere disclosure of the possibility of deception does not.
 - Experience with deception makes people more apprehensive of evaluation
 - The extent to which deception experience transfers appears to depend on the similarity between past and present situation.

Other costs of deception

- “Subjects were told point blank that they were in a conformity experiment. Quite a few speculated in the open-end question that the experiment was really a study of distance perception and that the material about being a conformity experiment was put in to throw them off the track! This type of response indicates that psychologists are disbelieved even in those rare situations in which they are caught telling the truth.” (Gallo, Smith, & Mumford, 1973, p. 106)

Other costs of deception (2)

- Methodological innovations is slowed down if fast, inexpensive, and convenient methods such as deception are available.
- Example - Bardsley (2000)'s replication of Weimann (1994)

Other costs of deception (3)

- The puzzle of the missing reputational spillover effects, explained?
 - Schultz (1969, Table 1) reports that on average less than forty percent of subjects were from introductory psychology courses.
 - Sieber & Saks (1989; see also Vitelli, 1988) report that of 74% of psychology departments that reported having subject pools, 93% reported they recruited them from introductory courses.
- => Does this change in recruitment mode reflect an evolutionary drift that was driven by attempts to minimize the contaminating effects of deception? (We think so.)

Policy recommendation

- *For economists:*
Don't abandon the proscription of deception. It's a simple but efficient convention.
- *For psychologists:*
Do without deception whenever you can. If you think you can't ...
[an incentive compatible mechanism for the reduction, if not elimination, of deception]

Conclusion

- Methodological discussion is good for us.
- What is needed is informed judgement based on:
 - previous research or
 - if no previous research available, the do-it-both-ways rule
- The do-it-both-ways rule increases experimental control and replicability. Hence facilitates theorizing.
- Shift in psychology's default practice concerning deception from first-resort strategy to last-resort strategy.

Does Bounded Rationality need new methodology?

- Not really.
- We do have non-invasive methodology such as MouseLab (e.g., Johnson, Camerer et al. Johnson & Cachon, Costa-Gomez et al.) to understand how people acquire information and perform decision and game theoretic tasks such as backward induction and forward induction. Or don't perform, for that matter.
- This methodology needs to be employed more widely to understand, e.g., variations in the costs of making decisions and judgements. Or the impact of emotions.
- We do have somewhat more
invasive procedures such as the various brain-scan techniques. These are important in terms of understanding the neural correlates of behaviors.
- It'd be nice to have less invasive procedures that do the same thing.
- Apart from that, we do have all the methodology that we need. We just need to apply it more.