Analyse & Kritik 29/2007 (© Lucius & Lucius, Stuttgart) p. 86–96

Daniel Houser/Daniel Schunk/Erte Xiao

Combining Brain and Behavioral Data to Improve Econometric Policy Analysis*

> Human behaviour may be governed by rules, but it is possible that these rules simply encode preferences. [...] Many psychologists argue that behaviour is far too sensitive to context and affect to be usefully related to stable preferences. However, if there are underlying preferences, then even if the link from preferences to rules is quite noisy it may be possible to recover these preferences and use them to correctly evaluate economic policies, at least as an approximation that is good enough for government policy work.
>
> Daniel L. McFadden (Nobel Prize Lecture 2002)

Abstract: For an economist, ultimate goals of neuroeconomic research include improving economic policy analysis. One path toward this goal is to use neuroeconomic data to advance economic theory, and productive efforts have been made towards that end. Equally important, though less studied, is how neuroeconomics can provide quantitative evidence on policy, and in particular the way in which it might inform structural econometric inference. This paper is a first step in that direction. We suggest here that key forms of preference (or decision strategy) heterogeneity can be identified by brain imaging studies and, consequently, linked stochastically to observable individual characteristics. Then, recognizing that brain-imaging studies are substantially costly, we derive conditions under which the probabilistic link between observable characteristics and type, a quantity critical to policy analysis, can be estimated more precisely by combining data from traditional and brain-based decision studies.

0. Introduction

For an economist, ultimate goals of neuroeconomic research include improving economic policy analysis. One path toward this goal is to use neuroeconomic data to advance economic theory, and productive efforts have been made towards that end (see, e.g., Camerer et al. 2005). Equally important, though less studied, is how neuroeconomics can provide quantitative evidence on policy, and

* We received useful comments from seminar participants at Caltech, George Mason University and the University of Montreal

in particular the way in which it might inform structural econometric inference. This paper is a first step in that direction.

We suggest here that key forms of preference (or decision strategy) heterogeneity, increasingly recognized as important to structural econometric policy analysis (e.g., Houser et al. 2004), can be identified by brain imaging studies and, consequently, linked stochastically to observable individual characteristics. Then, recognizing that brain-imaging studies are usually quite costly, we derive conditions under which probabilistic links between observable characteristics and type, the quantities critical for policy analysis, can be estimated more precisely by combining data from traditional and brain-based decision studies.

Recognizing heterogeneity in underlying individual characteristics or traits, i.e. different types of people, is important for social scientists. Discovering individual characteristics that are related to the enormous heterogeneity observable in human decision behavior will help social scientists to create better models of human behavior and, thereby, to evaluate the reaction of a population to economic policies. Identifying these types empirically is a challenging task, and it involves structural estimation.

Structural estimation refers to procedures by which one draws inferences about an economy's policy-invariant primitives such as preferences, endowments and technology. Structural estimation gained importance after Robert Lucas drew attention to the fact that individual economic decision patterns can change when policy changes, and that new individual decision patterns can disrupt historic relationships among macroeconomic quantities including consumption, savings and employment (see, e.g., Lucas 1977). Thus, he argued that econometric procedures focused on determining historic (partial) correlations among economic variables were potentially highly misleading approaches to policy analysis. On the other hand, if one were to focus on drawing inferences about policy-invariant characteristics of the economy, particularly preferences, then compelling policy analyses would in principle become possible.

The economic primitives relevant to a structural analysis are typically implied by economic theory and often represented by a finite parameter vector. As a practical matter, the usual goal of structural estimation is to value a theory's parameters in a defensible way. Neuroeconomics provides evidence on the appropriate specification and parameterization of individual preference primitives that enter economic theory, and also contributes to our ability to assign values to those parameters. In particular, neuroeconomic studies can inform the nature and number of individual 'types'.

Recently, econometrics has been developed to shed light on the nature and number of types (see, e.g., Houser et al. 2005 or Houser et al. 2004 and cites therein). The idea underlying them is to characterize relationships among individuals' information and decisions, and to determine the nature and number of clusters of similar relational patterns. These procedures have been used with success on data generated in controlled laboratory environments. However, whether types discovered in the laboratory games reveal policy-relevant preferences that can be extrapolated to naturally occurring environments, and whether existing econometric procedures can be used successfully to infer types in non-laboratory populations, both remain open questions.

Fortunately, neuroeconomic research holds the promise of making the nature and number of policy-relevant types, or preferences, in a population observable. In the best case, brain-data will reveal both the number of different utility functions in a population, as well as each function's shape and properties. Indeed, there exists evidence that types are in fact temporally persistent (see, e.g., Kurzban and Houser 2005) and that brain data can reveal deep aspects of utility functions (see, e.g., McCabe et al. 2001 for an fMRI study on human cooperation or Platt/Glimcher 1999 for single-cell neural evidence on monkey preferences).

In this paper we presume that neuroeconomic evidence has revealed the nature and number of utility functions, or types, in a population. Once in hand, economic theorists can use them to create compelling models for economic policy analysis. However, left unknown are (i) the population frequency of each type; and (ii) any links between an individual's type and their observable characteristics. Clearly, both of these are critical to know in order to make optimal policy decisions.

To estimate (i) and (ii) one could run very many imaging experiments. Unfortunately, such experiments are usually extremely costly both in terms of time and money. This often implies severe limits on the number of individual observations that can be feasibly collected. Consequently, inferences regarding type frequencies will likely be highly imprecise, both overall and within policy-relevant subsets of the population. This could leave the resulting policy analysis of questionable reliability. We are thus led to the central question of our paper: under what conditions can one use traditional (and less expensive) decision studies to improve inference regarding the (perhaps conditional) frequencies of the fundamental types derived from brain studies?

Figure 1 summarizes our discussion to this point. The top horizontal line represents the canonical economic policy-analysis problem: draw inferences about the stochastic relationship $p(d_{t+1}|x_t)$ between contemporaneously observable individual characteristics x_t (e.g., demographic background) and future decisions d_{t+1}. An unobservable factor influencing this relationship is a person's latent "type" k. Evidence suggests there are a finite number of types (see, e.g., Kurzban/Houser 2005 and cites therein), and current evidence suggests neuroeconomic studies will ultimately identify a finite universe Ω from which each person's preference is drawn. Thus, one can write

$$p(d_{t+1}|x_t) = \sum_{k \in \Omega} p(d_{t+1}|k, x_t)p(k|x_t).$$

Consequently, the policy-relevant prediction of interest is related to the conditional type probabilities $p(k|x_t)$. These conditional probabilities are not likely well-estimable from imaging data alone. In this paper we discuss conditions under which one can combine brain and behavioral data in order to improve the precision of those conditional probability estimates.

We develop our argument by sketching a generic parametric policy analysis framework within which 'type' enters in a meaningful way. We then discuss

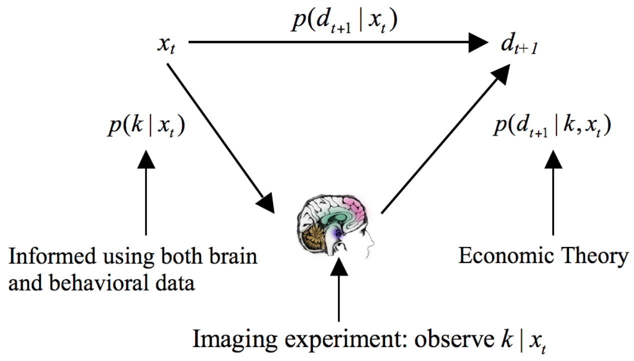

Figure 1: *Combining brain and behavioral data to make economic predictions. The canonical economic prediction problem is to draw inferences about the stochastic relationship* $p(d_{t+1}|x_t)$ *between contemporaneously observable individual characteristics* x_t *(e.g., demographic and educational background) and future decisions* d_{t+1}. *An unobservable factor influencing this relationship is a person's latent "type"* k. *Differences in type reflect differences in preferences or decision strategies that lead to different decisions among observationally identical individuals. Note that* $p(d_{t+1}|x_t) = \sum_k p(d_{t+1}|k, x_t) = p(k|x_t)$. *Hence, the behavioral prediction of interest is related to the conditional type probabilities* $p(k|x_t)$. *We develop the argument that these probabilities can be estimated using brain-imaging data, and derive conditions under which combining brain and behavioral data can improve the precision of those estimates.*

structural estimation of that model. In particular, we derive a condition that must be satisfied in order for the precision of structural inference regarding types to be improved by combining data from both behavioral and brain-imaging studies. For example, if the relevant types can be perfectly revealed by a behavioral study alone, then the condition is satisfied.[1]

This paper is in three sections. The next describes our framework formally. The second section offers simulation evidence on the value of our approach. The third section offers concluding remarks.

1. Neuroeconomics, Theory and Policy Analysis

1.1 Neuroconomics Informs Theory and Types

Advocates of neuroeconomics frequently point to its promise in informing new economic theory. For example, Colin Camerer writes: "Neuroeconomics is a subfield of behavioral economics which uses empirical evidence of limits on com-

[1] In this case one would still require brain imaging data in order to discover the types, and to discover that particular decisions do in fact perfectly reveal the preference types.

putation, willpower and greed to inspire new theories." (Camerer 2007, C26)[2] That this can and will occur does not seem controversial. Neuroeconomic research provides new data to correlate with economic decisions. Indeed, previous neuroeconomic research has correlated economic decisions with, for example, emotion expression (Xiao/Houser 2005), pupil dilation (e.g., Wang et al. 2006), hormone levels (e.g., Kosfeld et al. 2005) and fMRI brain activation patterns (e.g., McCabe et al. 2001). New economic theory will emerge to explain these novel correlations.

The connection between new theory and new data is especially clear in the case of brain activation studies. Brain studies promise to discover particular neural circuitries that implement particular economic behaviors. Once identified, and when combined with prior information regarding various areas of the brain, one has additional evidence to distinguish between competing economic theories. Moreover, because different neural circuitries implement different preferences, individual preference 'types' are observable.

1.2 Types and Policy Analysis

The goal of policy is to affect behavior. To be precise, let $d_{t+1} \in \{0, 1\}$ describe a policy relevant binary choice behavior. This could include, for example, whether one cooperates or not, shirks or not, or trusts or not. Broadly speaking, policy makers are interested in designing an institution $\zeta \in Z$ to promote a particular behavior, say cooperation, within an economic environment. Available to the policy maker, as well as to the econometrician helping to optimize the institution, is the population's vector of characteristics $x \in X$ as well as relevant information about the environment $e \in E$. The institution design problem is thus $\max_{\zeta \in Z} W(d(\zeta)|x, e)$, where W represents the policy maker's social welfare function, and $d(\zeta)$ denotes the dependence of the population's welfare-relevant choices on the institution.

Consider the case of a single decision between two alternatives, and assume the social welfare function is monotonically increasing in the (weighted) mean number of people who make the socially preferred decision. Then, if $\overline{\omega}_i$ represents the welfare-weight assigned by the policy maker to person $i \in I$, the policy maker's goal can be expressed as follows:

$$\max_{\zeta \in Z} \sum_{i \in I} \overline{\omega}_i p(d_{i,t+1} = 1 | x_{it}, e, \zeta) \tag{1}$$

Thus, in order to conduct policy analysis one must obtain estimates of the probabilities that appear in (1). This is generally the case for structural economic analysis. One obtains probabilistic predictions with respect to the way changes in policy affect a population's future decisions. As a result, policy analysis is improved if these probabilities are more precisely estimated.

To develop a framework for estimating these probabilities, focus now on a given environment e and a particular institutional structure ζ. Then, omitting

[2] See also Mullainathan/Thaler 2000; Camerer et al. 2005; Fehr/Camerer 2000.

this notation for convenience, the goal is to draw inferences about $p(d_{i,t+1} = 1|x_{it})$ for each person $i \in I$ of the population with observable characteristics x_{it} at time period t. Note that the goal is to assign a probability estimate to a decision that will be made in the future (period $t+1$) based on time period t observable characteristics.

In general, this probability depends on both observable and unobservable characteristics. Let c_{it} denote an individual's unobservable characteristics, and let s_{it} be the vector comprised of characteristics both observable and unobservable (to the econometrician). Separately, let k denote latent type, and assume probabilities can vary systematically with type. Finally, suppose that decisions have a stochastic component η. Thus, one is interested in the period $t+1$ decision of person i who is type k with (observed and unobserved) time t characteristics $s_{it} = x_{it} \cup c_{it}$. Denote this decision by $d^k_{i,t+1}(s_{it}, \eta_{i,t+1})$.

Next, let $p^k(\eta_{i,t+1}|s_{it})$ represent the conditional density of type *k's* error component, and $p^k(d^k_{i,t+1}|s_{it})$ the conditional density of the future decision of interest. Then,

$$p^k(d^k_{i,t+1} = 1|s_{it}) = p^k(d^k_{i,t+1}(s_{it}, \eta) = 1|s_{it}). \tag{2}$$

The typical goal of structural estimation is to inform the probabilities appearing in (2).

1.3 A Structural Model

Structural modeling involves placing restrictions on the way choice probabilities are affected by the elements entering (2). These restrictions are usually implied by theory. There are innumerable restrictions that theory might place on these choice probabilities. A weak restriction useful for our purposes, and that has been suggested by previous brain studies, is that types determine probability bounds. For example, McCabe et al. (2001) found that subjects showing significant activation in the medial pre-frontal cortex during a certain decision experiment were significantly more likely to cooperate than those who did not. Also, data reported by Sanfey et al. (2003) suggest that the probability of rejecting an offer in an ultimatum game is above 60% if the right insula exhibits "strong" activation and lower than 60% otherwise.

We interpret data such as these to imply that a person's type has been revealed. For example, in the Sanfey et al. (2003) study a person with significant activation is revealed to be a 'punisher' type, while those will less activation are revealed to be 'non-punisher' types. Thus, a scanning study allows one to determine from subject's brain-activation patterns whether they are one type or another. Moreover, one expects that a 'minimal' structural theory of types would assign different probability bounds to choices made by different people of different types. These bounds can be expressed as:

$$\begin{aligned} &(\forall s)p(d^1_{i,t+1} = 1|s_{it}) \geq \pi \qquad (3) \\ &(\forall s)p(d^2_{i,t+1} = 1|s_{it}) < \pi \end{aligned}$$

where $\pi > 0$ is implied by a theory that connects brain activation patterns to types, and types to probabilistic, policy-relevant decisions.

Within this framework, the goal is to predict a person's future decision given her observable characteristics and her behavioral type as revealed by their brain activation patterns.[3] Following Manski (1990), one way to do this is as follows.

First, from (3) it follows immediately that $p(d^1 = 1|x) \geq \pi$ and $p(d^2 = 1|x) < \pi$ where the time and person subscripts have been dropped for clarity. The reason is:

$$p(d^k = 1|x) = \int p(d^k = 1|s) dP^k(s|x) \Rightarrow$$
$$p(d^1 = 1|x) = \int p(d^1 = 1|s) dP^1(s|x) \geq \pi \int dP(s) \geq \pi; \text{ and}$$
$$p(d^2 = 1|x) = \int p(d^2 = 1|s) dP^2(s|x) < \pi \int dP(s) < \pi.$$

Note that these bounds do not vary with individual characteristics. Instead, brain activations imply a type, which acts as a summary statistic for a person's expected decision patterns. Thus, from information on a person's type, derived from an analysis of her activation patterns, one can obtain substantive bounds on the probability with which a person will make a decision based only on characteristics that are observable to an econometrician. Note further that this is a testable restriction: data on types and decisions actually made can refute the bounds implied by this specification.

Of course, one does not have information on type for each individual in a relevant population. However, one does have information on characteristics x from a significant subset of the population. From this one can derive the following restriction on $p(d = 1|x)$. Focusing on a single individual, and recalling that k denotes a person's type, first note:

$$p(d = 1|x) = p(d^1 = 1|x)p(k = 1|x) + p(d^2 = 1|x)p(k = 2|x) \tag{4}$$

and from this it follows that

$$\pi\, p(k = 1|x) \leq p(d = 1|x) \leq \pi\, p(k = 2|x) + p(k = 1|x) \tag{5}$$

Thus, the probability that a person with characteristics x makes the decision of interest is bounded from above and below by quantities that depend on his or her observable characteristics x as well as on the theory that determines π. Because theory determines π these bounds are useful—and policy analysis substantively informed—to the extent that the relevant conditional (on observed characteristics) type probabilities can be estimated. To do this one could run many imaging experiments. However, such an approach would likely be prohibitively costly both in terms of time and money. Fortunately, in some cases

[3] We assume throughout that type is determined with certainty by the brain-imaging patterns. Extending our argument to accommodate imperfectly determined types is left to future research.

the policy-relevant conditional type-probabilities can be informed by incorporating data from traditional economic decision experiments.

1.4 Behavioral experiments can inform conditional type probabilities

We argued above that individual differences are identified by patterns of brain activation in neuroeconomic brain-imaging experiments. In this section we continue to focus on the case of two types, and derive conditions under which decisions in traditional economic experiments can be combined with decisions in an imaging experiment to improve inference about the policy-relevant conditional type probabilities.

Denote by Y the action set of the decision experiment, with typical element $y \in Y$. Whether actions in the decision experiment inform the conditional choice probabilities clearly depends on how those actions are related to brain activation. To develop this idea formally, first observe:

$$p(k = j|x) = \frac{p(k = j|y, x)}{p(x|k = j, x)} p(y|x), j = 1, 2. \tag{6}$$

Both the numerator and the denominator of the ratio that appears in (6) require one to know both an individual's type as well as the decisions made in the experiment. Thus, neuroeconomic brain-imaging experiments are needed in order to draw inferences about that ratio. In contrast, the final term in (6) is simply the conditional density of actions in an economic decision experiment, a quantity which can be estimated absent imaging data.

It is useful to express (6) in log form. Doing so yields:

$$\log p(k) = \log p(y) + \log p(k|y) - \log p(y|k) \tag{7}$$

where for convenience the conditioning variable x has been suppressed, and it is understood that, for example, $k = 1$. The term on the left hand side can thus be estimated by obtaining estimates of the three quantities on the right hand side. Each of these quantities is identified, and in typical cases estimable using simple frequency procedures. Let $\hat{p}_y$ denote the estimate of $\log p(y)$ and similarly for the other quantities. Then, the variance of the estimator of $\log p(k)$ can be expressed as follows.

$$\text{var } \hat{p}_k = \text{var } \hat{p}_y + \text{var } \hat{p}_{k|y} + \text{var } \hat{p}_{y|k} + 2\text{cov}(\hat{p}_y, \hat{p}_{k|y}) - 2\text{cov}(\hat{p}_y, \hat{p}_{y|k}) \tag{8}$$

$$-2\text{cov}(\hat{p}_{y|k}, \hat{p}_{k|y}) = \sigma_y^2 + \sigma_{k|y}^2 + \sigma_{y|k}^2 + 2\sigma_{y,k|y} - 2\sigma_{y,y|k} - 2\sigma_{y|k,k|y}.$$

Now, simple frequency procedures estimate $\hat{p}_y$ consistently. Consequently, we have from (8) that, as the number of observations becomes large, $\sigma_y^2 \to 0$, $\sigma_{y,k|y} \to 0$ and $\sigma_{y,y|k} \to 0$. Combining this fact with (8) implies that behavioral experiments can be combined with imaging experiments to improve the precision of the estimated type probabilities when

$$\sigma_y^2 > 2(\sigma_{y,y|k} - \sigma_{y,k|y}). \tag{9}$$

We offer two simulation exercises to help clarify when this condition holds.

2. Simulations

This section reports the results of two simulation experiments we ran to help provide intuition for cases where condition (9) will hold, and when it will not. Consider first the case where behavioral experiments should clearly not be able to provide additional information about a person's type. From (6), one such case is when action in the behavioral experiment is independent of a person's type. In this case $p(y|k=1,x) = p(y|x)$ so that behavioral experiments cannot inform conditional probabilities. Hence, condition (9) should not hold in this case.

To investigate this, we imagine an imaging experiment where subjects can choose either 'right' (say, cooperate) or 'left' (say, defect). The way a subject thinks about this problem reveals their type in an imaging experiment, but choice frequencies are nevertheless independent of this type. To capture this idea we simulated data using the following process:

$$p(k=1) = 0.5$$

$$p(y = \text{“right”}|k=1) = 0.5$$
$$p(y = \text{“right”}|k=0) = 0.5$$

Thus, marginal type probabilities are 0.5, the marginal probability of choosing 'right' is 0.5, and whether a person chooses 'right' is formally independent of her type.

We simulated 3000 samples of size N=25 each. The variances in (7) are estimated as the variance of the sample of 3000 means. Thus, for example:

$$\begin{aligned}
\hat{p}_{jy} &= \sum_{i=1}^{25} \chi \qquad (i \text{ in simulation } j \text{ chooses ‘right’})/25, \\
&\qquad\qquad \text{where } \chi(z) = 1 \text{ if ‘z’ is true, and zero otherwise,} \\
\bar{\hat{p}}_y &= \sum_{j=1}^{3000} \hat{p}_{jy} / \, 3000 \\
\sigma_y^2 &= \sum_{j=1}^{3000} \left(\hat{p}_{jy} - \bar{\hat{p}}_y\right)^2 / \, 2999
\end{aligned}$$

and similarly for the other relevant terms.

Doing this reveals that $\hat{\sigma}_y^2 = 0.04$ while the right hand side of (9) is 0.09. Thus, as must be true in this case, condition (9) does not hold. That is, collecting behavioral data alone cannot inform the desired conditional type probabilities.

Intuitively, behavioral experiments alone are expected to inform the relevant conditional type probabilities when there is significant dependence between a person's type and her decision in the behavioral task. To address this possibility, we simulated new data using the following structure.

$$p(k = 1) = 0.5$$

$$p(y = \text{'right'}|k = 1) = 0.85$$

$$p(y = \text{'right'}|k = 0) = 0.15$$

Thus, the marginal type probabilities remain the same, and the overall frequency of "right" choices remains the same at 50%, but the probability that a given individual chooses right varies substantially with her type. Following the same procedures as described for the first simulation again gives $\hat{\sigma}_y^2 = 0.04$ but the right hand side falls to 0.03. It follows that in this case the statistical precision of the estimated conditional probabilities is improved by up to 25% with imaging data alone.

3. Conclusion

Economists have ultimate interest in designing institutions to promote the efficient exchange and allocation of scarce resources. Some have questioned how or whether neuroeconomics fits into the boundaries of this traditional economic paradigm. This paper offers one response to that question. We pointed out that modern institution design recognizes the importance of individual differences in preferences when forming policy advice, and argued that neuroeconomic research provides important information on the nature of individual differences. Furthermore, we demonstrated that this information is valuable, and can be exploited to improve policy analysis, even when only a small fraction of the relevant population has taken part in an imaging study. In particular, we derived a condition under which traditional 'mindless' data can be combined with brain imaging data to improve the precision of structural estimates of policy-relevant conditional type probabilities.

Bibliography

Camerer, C./G. Loewenstein/D. Prelec (2005), Neuroeconomics: How Neuroscience Can Inform Economics, in: *Journal of Economic Literature 43*, 9–64

— (2007), Neuroeconomics: Using Neuroscience to Make Economic Predictions, in: *Economic Journal, 117(519)*, C26–C42

Fehr, E./C. Camerer (2007), *Social Neuroeconomics: The Neural Circuitry of Social Preferences*, Forthcoming, Trends in Cognitive Science

Houser, D./M. Keane/K. McCabe (2004), Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm, in: *Econometrica 72(3)*, 781–822

—/A. Bechara/M. Keane/K. McCabe/V. Smith (2005), Identifying Individual Differences: An Algorithm with Application to Phineas Gage, in: *Games and Economic*

Behavior 52, 373–355

Kosfeld, M./M., Heinrichs/P.J. Zak/U. Fischbacher/E. Fehr (2005), Oxytocin Increases Trust in Humans, in: *Nature 435(7042)*, 673–676

Lucas, R.E. (1977), Understanding Business Cycles, in: *Carnegie-Rochester Conference Series on Public Policy 5*, 7–29

McCabe, K./D. Houser/L. Ryan/V. Smith/T. Trouard (2001), A Functional Imaging Study of Cooperation in Two-person Reciprocal Exchange, in: *Proceedings of the National Academy of Science 98(20)*, 11832–11835

Manski, C. (1990), The Use of Intentions Data to Predict Behavior: A Best Case Analysis, in: *Journal of the American Statistical Association 85*, 934–940

Mullainathan, S./R. Thaler (2000), *Behavioral Economics: Entry in International Encyclopedia of the Social and Behavioral Sciences*, Cambridge

Platt, M.L./P.W. Glimcher (1999), Neural Correlates of Decision Variables in Parietal Cortex, in: *Nature 400*, 233–238

Sanfey, A.G./J.K. Rilling/J.A. Aronson/L.E. Nystrom/J.D. Cohen (2003), The Neural Basis of Economic Decision-making in the Ultimatum Game, in: *Science 300*, 1755–1758

Xiao, E./D. Houser (2005), Emotion Expression in Human Punishment Behavior, in: *Proceedings of the National Academy of Sciences 102(20)*, 7398–7401

Wang, J.T.-Y./M. Spezio/C.F. Camerer (2006), *Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth-telling and Deception in Biased Transmission Games*, Caltech working paper