

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2013/1**



**Behavioral Law
and Economics:
Empirical Methods**

Christoph Engel



MAX PLANCK SOCIETY



Behavioral Law and Economics: Empirical Methods

Christoph Engel

January 2013

Behavioral Law and Economics: Empirical Methods*

Christoph Engel

Abstract

Originally, behavioral law and economics was an exercise in exploring the implications of key findings from behavioral economics (and psychology) for the analysis and reform of legal institutions. Yet as the new discipline matures, it increasingly replaces foreign evidence by fresh evidence, directly targeted to the legal research question. This chapter surveys the key methods: field evidence, survey data, vignette and lab experiment, discusses their pros and cons, illustrates them with key publications, and concludes with methodological paths for future development. It quantifies statements with descriptive statistics about the 77 behavioral papers that have been published in the *Journal of Empirical Legal Studies* since its foundation until the end of 2012.

JEL: C01, C83, C91, D02, D03, K00

Keywords: behavioral law and economics, law and psychology, criminology, field data, survey data, vignette, lab experiment

* Helpful comments by the editors, Angela Dorrough and Pascal Langenbach are gratefully acknowledged

1. The Landscape

Eventually, all law is behavioral. In a forward-looking perspective, this is obvious. In this perspective, law is a governance tool (see only Posner 2011). All the law is potentially able to achieve is a change in human behavior. Of course, not all lawyers agree with this mission statement. In a backward-looking perspective, the critical issue is attribution. The minimum requirement for attribution is possible cause. A person has possibly acted in some way, and this has possibly led to detriment. Whether the person has actually acted is a behavioral question. Usually the law is not content with mere causation and requires intent or negligence. That makes the behavioral dimension even more prominent. Of course the person may be a legal person, e.g. a firm. Then the additional complexity resulting from the presence of a corporate actor must be handled (Engel 2010). Still the core of the matter is how the law should react to some course of behavior. Finally, in a deontological perspective, the law emanates from and impacts on moral intuitions prevalent in society. Whether individuals truly hold a purported moral intuition, and how these intuitions and explicit legal rules interact, is another behavioral issue (Zamir and Medina 2011).

In two areas of law, a behavioral perspective has a long tradition. Law and psychology scholars for long have studied psychology in the courtroom. Criminologists have long tried to understand why there is crime and how it can be mitigated by criminal law. Behavioral law and economics is a more recent phenomenon (see the programmatic book by Sunstein 2000). As law and economics in general, it adopts an individualistic perspective. Legal rules are understood (and often explicitly modeled) as changes in the opportunity structure. The overall research question is explaining actions of individuals by their reaction to this opportunity structure. Behavioral law and economics deviates from traditional law and economics by the assumptions about the driving forces of individual action. In the standard framework, individuals have well-defined and well-behaved preferences, they know everybody else's preferences, and they dispose of unlimited cognitive abilities. In line with the behavioral approach to all economics, behavioral law and economics relaxes these assumptions. It allows for richer utility functions, and in particular for social preferences. Actors are no longer assumed to exclusively care about their own well-being. Behavioral law and economics is also open to a panoply of cognitive effects, and investigates in which ways legal institutions reflect or mitigate them. Behavioral law and economics also adopts a broader topical scope than the two older behavioral traditions. It focuses on the behavioral foundations of all law, and in particular of private and public law.

Most behavioral lawyers are consumers of scientific evidence. They for instance capitalize on the heuristics and biases literature in cognitive psychology (for a summary account Kahneman and Tversky 2000) and investigate its implications for the interpretation and the design of the law. Most of these contributions come out in the law reviews (e.g. Rachlinski 2011). Sometimes a broader theme leads to a monograph. I have for instance argued that, taking the existing behavioral evidence into account, human behavior seems utterly unpredictable. This points to a neglected purpose of legal intervention. Rather than taming socially undesirable

motives, through institutional intervention the law makes behavior predictable and thereby social interaction meaningful (Engel 2005).

In many dimensions, the existing behavioral evidence is rich and differentiated. Lawyers have no reason to reinvent the wheel. Yet not so rarely a lawyer is unable to spot the evidence she needs for her normative business. There are two main reasons why a neighboring discipline has not delivered. The lawyer may want to know whether a general effect also holds under the specific conditions the law aims to address. Or the specific behavioral effect the lawyer suspects to be critical has just not been on the screen of psychologists or economists. In such situations, legal scholars would ideally wish to generate fresh behavioral evidence. This of course requires expertise in empirical methods, or collaboration with colleagues who have this training. It is a happy coincidence that the empirical legal movement is growing so rapidly these days. Many believe that the movement has chiefly been ignited by the availability of big data. By far not all empirical law is behavioral. But a substantial fraction is, and it benefits from mounting interest in the empirical foundations of doctrinal or legal policy argument.

Sometimes, trained lawyers publish in the psychology (e.g. Simon et al. 2004) or in the economics journals (e.g. Zeiler and Plott 2005). Sometimes, fresh behavioral evidence is the core of a law review article (e.g. Buccafusco and Sprigman 2011). Yet the typical outlet for lawyers' efforts at generating new behavioral evidence is the peer reviewed legal journals. If the paper is written in the psychological paradigm, and uses the methods prevalent in psychology, there are two specialized journals, *Psychology*, *Public Policy & Law*, and *Law & Human Behavior*. The former has a slightly broader focus and publishes papers from all subfields of law. It is particularly interested in contributions with direct policy relevance. The latter journal specializes in the two intersections between law and behavioral research with the longest tradition: criminology and forensic psychology. In principle, all of the many criminology journals are open to behavioral evidence on crime and criminal law. This in particular holds for the flagship journal *Criminology*. Two journals are particularly relevant since they focus on one empirical methodology, the *Journal of Experimental Criminology* and the *Journal of Quantitative Criminology*.

These specialized outlets notwithstanding, scientific empirical evidence is closely tied to law and economics. All the relevant peer-reviewed journals also publish behavioral evidence, the *Journal of Law and Economics*, the *Journal of Law, Economics and Organization*, the *Journal of Legal Studies* the *American Law and Economics Review*, the *International Review of Law and Economics* and the *Review of Law and Economics*. The same holds for the newly founded *Journal of Legal Analysis*. Yet in all these journals, empirical contributions compete with theory and policy papers. This is different with the *Journal of Empirical Legal Studies* (hereinafter *JELS*). The journal is entirely devoted to empirical contributions. In the following paragraphs, I focus on this journal since this generates the most accurate portray of the emerging discipline.

The journal has been founded in 2004. In the first nine years of its existence, it has published a total of 227 articles. 77 can be classified as behavioral.¹ This makes for 33.92% of all publications. Figure 1 shows that, a slight drop in 2012 notwithstanding, the number of behavioral publications has been steadily growing over time, both in absolute terms and relative to the total number of publications in the respective year.

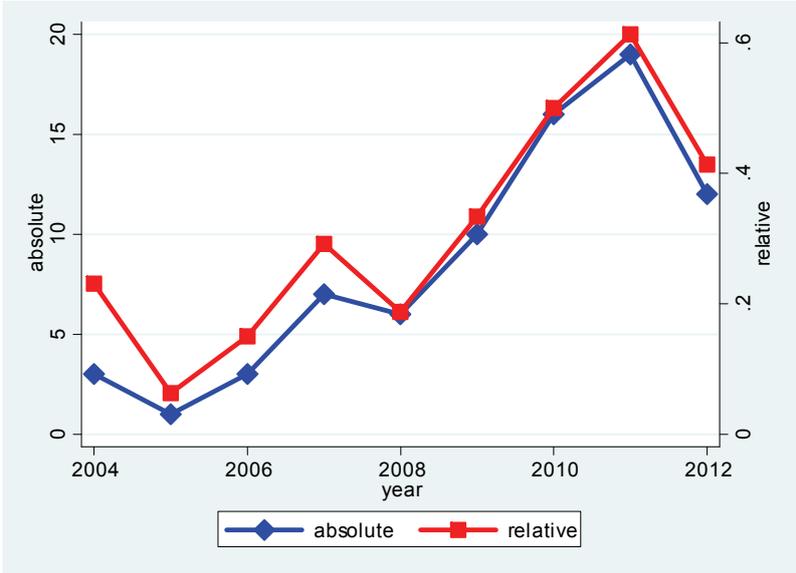


Figure 1
Behavioural Evidence in the *Journal of Empirical Legal Studies*

32 behavioral papers deal with a question of private law. 27 cover an issue from criminal law. Public law has only attracted 5 publications. The remaining 13 papers are not focused on one subdiscipline of law in particular.

Empirical scholars have been much more interested in motivation (57 publications), rather than cognition (20 papers).

Publications are strongly US-centric. 67 papers either address a question from US law, or they use US data or, in an experiment, US subjects. There are three experimental papers from Germany, two papers from the UK, and one from Australia, Israel, and Taiwan each. Finally two papers compare different jurisdictions.

In the following, I will use the *JELS* data to describe and analyze the methodological variance (section 2). By way of illustration, I will also hint to publications from the remaining peer-reviewed legal journals. In conclusion, I will sketch paths for future methodological development (section 3).

1 I have coded all papers as behavioral that explore an aspect of motivation or cognition, from whatever conceptual or methodological perspective. The dataset coding these publications is available upon request.

2. Competing Methodologies

a) Choice of Method

Method matters. Let me illustrate this with one of the prominent claims of law and economics theory. Traditionally, tort has been constructed as a technology for restoration. The tortfeasor has intruded into the victim's sphere without entitlement, and has caused harm. The victim sues the tortfeasor for redress. The court ruling is meant to make the victim whole. Law and economics scholars object: this construction neglects that rational would-be tortfeasors anticipate the intervention and adjust their behavior. In this forward-looking perspective, tort liability deters socially undesired behavior, much like a criminal sanction or the intervention of a public authority. Quite some lawyers have been skeptical whether liability is indeed a powerful deterrent. There have been multiple empirical attempts to measure the effect, and the evidence has been mixed (Schwartz 1994).

Recently, three different empirical studies have tried to settle the issue. One study worked with field data. If tort liability is more severe, it should induce would-be tortfeasors to be more careful. This should reduce the overall level of harm. The study used an indirect approach to measure the effect. It tested the following hypothesis: in those US states with stricter rules on medical malpractice, newborns should be in better health. The study did not find a significant association between the two (Yang et al. 2012).

The second study used the methodology prevalent in psychology. First year law students were exposed to a series of hypotheticals, all of which involved some form of illegal behavior. Between subjects, the authors manipulated the legal regime. One of those regimes was tort liability. Students were asked to indicate how likely they were to engage in one of these activities, given the legal regime in question. Tort liability had little effect. For many scenarios, the proclivity to engage in illegal activity was not significantly different from the condition where no legal rule was mentioned (Cardi et al. 2012).

The third study used the methodology developed in experimental economics, i.e. a lab experiment. Participants were exposed to a four person dilemma. If a participant behaved selfishly, she imposed damage on the remaining group members. The experiment manipulated the certainty and the severity of the right to claim redress. If redress was sufficiently certain, if it was sufficiently severe, and if the threat of compensation had a sufficiently high expected value, the dilemma did not disappear, but it no longer deteriorated over time. With this qualification, the experiment found a deterrent effect of tort (Eisenberg and Engel 2013).

Of course, these three studies differed by more than just methodology. It could well be that doctors are less sensitive to the severity of tort liability than students. It could be that professionals are less sensitive than ordinary people. It could be that doctors care less because they are insured. It could be that severity matters for intentional tort but not for negligent tort. It could be that severity matters if the tortuous act simultaneously harms multiple victims, but

not in a one-to-one relationship. Yet it could also be that one needs a method where individuals actually feel the pecuniary loss of paying damages to see the governance effect of tort.

Behavioral legal scholars have used a considerable plurality of empirical methods. Figure 2 shows that two methods are most prominent: field data and vignette studies. Currently, surveys are less visible in *JELS*. Lab experiments have always been least popular.

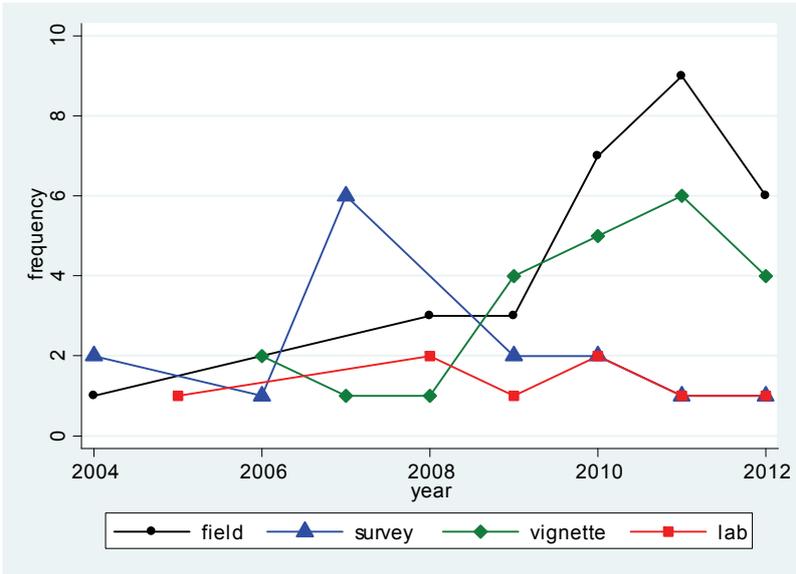


Figure 2
Empirical Methodologies in the *Journal of Empirical Legal Studies*

Any choice of empirical method is a tradeoff between external and internal validity. A result is externally valid if there is a good match between the data and the social phenomenon the researcher intends to explain or predict. Seemingly with observational data, external validity is beyond doubt. One directly studies the phenomenon one wants to explain. Yet behavioral researchers have no reason to expect natural laws. They may at best find typical patterns. Ideally they also gain a sense of relative frequency, and of robustness to contextual variation. It therefore would not be meaningful to hunt the one observation that disproves the claim. Behavioral researchers take it for granted that such exceptions exist. They are content with delineating the conditions under which an effect is normally present.

External validity would still not be an issue if researchers could simply observe the total population. They could then map out the framework conditions of the effect simply by studying it under all possible conditions. Yet behavioral research next to never observes a total population. Probably the empirical work on decision-making in the US Supreme Court comes closest. But even there many steps of the decision-making process remain confidential. More importantly researchers want to predict future decisions by the court. Even if the composition of the court remains stable, this still requires extrapolation from the past to the future. Normally, behavioral researchers also want to extrapolate from the sample they have observed to the population they want to explain.

For such extrapolation to be legitimate, researchers must be sufficiently confident that the effect they have observed is characteristic for the population. Assessing this confidence is the purpose of statistical analysis. Technically, the researcher compares some characteristic feature of her observation with some well-defined counterclaim. She for instance has expected that the requirement to justify administrative orders in writing leads to a greater willingness to abide by the order. Let's assume there are two otherwise similar administrative decisions. The law obliges the authority to give written reasons in the first domain, but not in the second. Let us further assume that the researcher finds sufficient qualitative evidence to support the claim that the number of appeals is a reasonable proxy for the willingness to abide by the order just because it is in force. On these assumptions, the researcher could count the number of appeals in both domains. Her statistic would inform her whether this difference is so large that it is very unlikely to result from random variation.

In research practice, this test is entrusted to a statistical package. Researchers usually only report that they have found a significant effect. According to the convention in the social sciences, this statement is warranted if the probability that the difference results from randomness is below 5%. What easily gets lost, though, is that this procedure requires an explicit *ex ante* hypothesis. The hypothesis is surprisingly often not made explicit. It was missing in 29 of the 77 empirical behavioral papers in *JELS*.

Even if significance is established, this only shows that one phenomenon is associated with another. All one knows is correlation. Typically this is not enough for the normative legal research question. There are three classic topics. Legal scholars want to understand whether there is reason for legal intervention. They want to know whether some legal rule is likely to improve the situation. And they want to assess the quality of some process for rule generation or rule application. All these research questions require the separation of cause and effect. Because consumers are misled by some marketing technique, this technique should be banned. Because individuals understand probability information better if it is presented in the form of natural frequencies, a rule that obliges insurance companies to use this format helps individuals make better choices. Because individual consumers are very unlikely to sue a company for the use of a detrimental standard form contract, class action advances consumer rights.

Empirical methods are on a continuum. Field evidence has the key advantage that one directly studies the phenomenon one wants to understand. Yet with field data, identification is notoriously problematic. The more the researcher takes identification seriously, the more she is at the mercy of unanticipated natural variation. On the other end of the spectrum are lab experiments. They artificially generate an environment, and randomly expose participants to different regimes. If the experiment is well designed, these regimes differ by just one feature. If there is a treatment effect, it must result from this manipulation. Causation is not an issue. But experimenters pay a price. Of necessity, what they study is only analogous to what they want to understand. External validity almost always is an issue. Surveys are one step more artificial than field data. They are handed out to those whose behavior one wants to understand. But the

set of questions is designed, not naturally occurring. Vignette studies share many features of lab experiments. They differ by being hypothetical. Participants are asked to assess scenarios, rather than make decisions that matter. These scenarios typically tell a real-life story, and thereby admit considerably more context. Surveys and vignettes are thus in the middle between both extremes. In the following, the power and the limitations of these competing empirical methods are illustrated.

b) Field Evidence

Behavioral researchers are interested in motivation and cognition. These are difficult to observe directly. It is therefore not obvious how observational data helps answer behavioral research questions. Three very different publications illustrate options.

A first study exploits the fact that juries have to take a joint decision. In preparation, juries deliberate. The state of Arizona considered allowing juries to discuss the evidence during trial. To learn more about the effects of such a reform, the Arizona Supreme Court sanctioned videotaping of randomly selected juries. Researchers exploited the opportunity to test whether jury members were influenced by information that had not been officially introduced into court procedure, like signals for a witness being wealthy. The study found that this was indeed the case, but that such “offstage” information had little impact on later jury deliberations about guilt (Rose et al. 2010).

Judges should treat all defendants on the merits of their cases. This norm implies that the defendant’s ethnicity should not matter per se. In principle, ethnic bias is difficult to identify since ethnicity is notoriously correlated with the merits of many cases. If one ethnic group is more violent than another, members of this group should be punished more often for violent crime. Yet the individual defendant should not be more likely to be convicted just because of his ethnic background. A study used the fact that, in Israel, over the weekend, for bailing suspects are randomly assigned to judges. It turned out that Jewish judges were more likely to bail Jewish than Arab suspects, while it was the other way round for Arab judges (Gazal-Ayal and Sulitzeanu-Kenan 2010).

Many diseases are not exclusively the result of factors beyond the individual’s control, like genetics or accident. Attempts at helping the needy may therefore have the counter-productive effect of reducing self-control. In principle, this offsetting effect is difficult to show since almost all diseases have multiple causes, at least potentially. A study uses legal reform as a source of variation. Some US states have made it mandatory that health insurance plans cover diabetes treatment. The study uses body mass index as a proxy for self-control. Arguably, if diabetics eat more, they increase diabetes problems. The study finds that the gap in the body mass index between diabetics and non-diabetics increased in states that had made coverage for diabetes mandatory (Klick and Stratmann 2007).

Establishing causality is difficult with observational data for two major reasons. Significant correlation may result from reverse causality. Because consumers know that the legal order cares, they stop being vigilant. Significant correlation may also result from an omitted variable that explains both the presumed cause and the presumed effect. The true cause for both the choice of the marketing technique and its effect is a lack of literacy in one group of consumers (classic Leamer 1983).

Econometricians have developed a whole panoply of techniques to solve such endogeneity problems (for an excellent introduction see Blundell and Costa Dias 2009). Essentially they all rest on some form of quasi random variation. Empirical legal scholarship on behavioral issues is very differently sensitive to this identification problem. 32 of 77 papers published in *JELS*, and 12 of the 29 studies working with observational data, do not deal with the identification problem at all. Those observational studies that address identification use many different approaches.

The most congenial approach to empirical legal studies is the difference in difference estimator. It is used by 6 of the 29 relevant publications in the *JELS* (a good example is Frakes 2012). Some legislator has changed a critical rule while another legislator has not. This permits identification if the two jurisdictions are sufficiently comparable and if the rule change was the only major difference in development between them. The researcher may then compare the change in the outcome variable before and after the rule change with the change in the same outcome variable in the jurisdiction where rules have not changed. If there is a significant difference between these two changes in outcomes, it is caused by the change in rules. This procedure has the advantage that it cleans the data from unobserved effects that are idiosyncratic to each jurisdiction. Since the procedure only looks at changes, idiosyncrasies are immaterial as long as they do not change over time, and to the extent that they do not interact with the rule change.

The classic response of econometricians to endogeneity problems is instrumentation. One tries to find an additional variable that is sufficiently correlated with the endogenous explanatory variable, but uncorrelated with the dependent variable. In the most straightforward application, in a first step one explains the potentially endogenous variable by the additional exogenous variable. In the equation explaining the dependent variable one replaces the endogenous variable by the predicted values from the first estimation step. Essentially one now explains the dependent variable by that portion of the independent variable that is cleaned from reverse causation or the omitted variable problem.

A single behavioral paper published in *JELS* uses instrumentation (Brinig and Garnett 2012). The paper investigates one facet of the relationship between religion and crime. The paper hypothesizes that the presence of a Catholic elementary school creates social capital which, in turn, reduces crime. Obviously there could be reverse causality. Catholic schools could be closed because the community deteriorates. Therefore correlation per se is not informative. The authors instrument school closure with irregularities in the parish. Arguably such irregu-

larities are not correlated with social capital in the respective community. They find the effect of religion they were expecting.

Occasionally, the law generates randomness for reasons unrelated with research. For instance, in some branches of the US Court of appeals, judges are randomly assigned to cases. This makes it possible to test for bias resulting from proximity of a judge to the political party of the president by whom she has been appointed. Such bias may indeed be established (Hall 2010).

c) Survey

Observational data leaves the phenomenon one wants to understand completely untouched. Surveys are slightly more intrusive. Participants are interviewed, maybe in writing or online, on their knowledge, understanding, attitude, judgement, or choices. Usually, one and the same participant answers a whole battery of questions. Typically questionnaires are handed out to members of that same population whose behavior one wants to understand. One such study was interested in inpatients falsely confessing or falsely pleading guilty (Redlich et al. 2010). To that end, interviews with detainees in six different institutions were conducted. A single institution significantly differed from the remaining institutions, so that the main findings could be replicated multiple times. Although participation was not random, this makes it very unlikely that the effects result from selecting atypical cases. Members of minorities, those with a longer criminal career, and those more severely mentally ill were significantly more likely to report that they had wrongly confessed or pled guilty. The paper discusses the limitations inherent in self-report. It does not address the identification problem: did participants plead guilty because they were mentally ill, or did they become mentally ill because they pleaded guilty?

d) Vignette

The standard method in law and psychology is a vignette study. Participants are presented with a hypothetical scenario, and they are asked how they would behave themselves were they in that situation, or how they would react were they to learn about such behavior. In jury research, this method is standard because, in principle, in the US researchers are not admitted to the jury box. Even if direct observation is not legally prohibited or technically impossible, researchers may prefer a vignette study since it eases identification. In a between subjects design, participants are randomly assigned to different versions of the vignette. Any treatment effect may then be traced back to the difference between the scenarios. In a within subjects design, every participant reacts to more than one (qualitatively different) vignette. If the reaction differs, this difference must result from having seen the earlier vignettes.

To illustrate, consider a study on the standard of proof in civil litigation (Zamir and Ritov 2012). In the US, the official standard is preponderance of the evidence. In the literature, this

standard is frequently translated into the posterior probability of the claim being well founded to be above 50%. Earlier evidence suggests that triers actually require a higher probability. The study proceeds in two steps. It first has separate groups of law students rate the persuasiveness of the plaintiff's case. Another group of law students is asked whether they would find for the plaintiff. Despite the fact that the mean rating of persuasiveness is above 50% for all three scenarios, a much smaller group declares that they would find for the plaintiff. The result replicates with professional lawyers. In the second step, a new group of students is handed out a questionnaire that explores potential explanations. The one explanation that stands out is what the authors call omission bias. Participants believe that judges dread responsibility if they find for the plaintiff despite the fact that the evidence is weak. Consequently participants believe that judges feel less responsibility if they erroneously dismiss the claim.

Vignette studies are experiments. The experimenter manipulates the scenarios. Participants are randomly assigned to different versions, or they see different versions over time. Sometimes participants are also alien to the situation the experimenter wants to explain. Not so rarely, participants are students, although the study wants to explain the behavior of the general public, or of legal officers. Many researchers see this as a limitation, and prefer giving the vignettes to a sample that is representative for the general public (12 of the 23 vignette studies published in *JELS*) or to legal officers (6 papers), usually citizens on jury duty.

e) Lab Experiment

The standard tool of experimental economists is the lab experiment. There are seven main differences, compared with a vignette study. (1) The main, if not the exclusive dependent variable is a choice. This choice is incentivized. What participants decide, and how they interpret the situation in preparation of their decision, directly matters for the payoff they receive. (2) Interest lies in abstract effects. To test for a hypothesized effect, the design is free from context. The typical design is a game. The benchmark solution is provided by the response of a person exclusively motivated by pecuniary gain, and in possession of unlimited cognitive abilities. Usually, this prediction is contrasted with an alternative hypothesis based on a richer utility function, or assuming less than perfect cognition. (3) Most economic experiments are interactive. Participants are not studied in isolation but as they interact. (4) Many economic experiments repeat a stage game multiple times. This is done in the interest of studying how effects unfold over time. (5) There is a culture in economic labs that forbids deception. This rule is meant to improve identification. Experimenters want to be sure that any effects indeed result from their manipulation, not from participants second-guessing how the experimenter tries to trick them today. (6) Economic experiments are usually completely computerized. They are run in a computer lab. Complete anonymity is guaranteed. Usually all communication is through choices. These precautions also aim at better identification. (7) Often, hypotheses are derived from a formal model. Actually, the experimental methodology is meant to

directly map formal economic theory, and game theory in particular (more discussion of the different experimental paradigms in Hertwig and Ortmann 2001).

Compared with the remaining empirical methods, lab experiments put most stress on internal validity. Predictions are as precise as possible, and as clearly grounded in explicit theory as possible. Observations are made as credible as possible. The design tries hard to exclude alternative explanations for the treatment effect, even those resulting from the construction of the situation. The price economic experiments pay for this rigor is less external validity. The policy problem that motivates the experiment is not directly visible. Economic experiments make a contribution to the policy discourse by isolating one effect. There is always reason to discuss whether alternative explanations or additional factors still support intervention. Yet by stripping the situation from all context, and by translating it into a naked incentive structure, one is able to study this driving force.

I illustrate the potential and the limitations of economic experiments for legal issues with one of my own papers (Engel and Kurschilgen 2011). German copyright law has a seemingly odd provision. If a work, say a film, turns out a blockbuster, those who have contributed may sue the producer and claim additional remuneration. We have translated this into a sequential two-person game. At the beginning of the game, both players received an equal endowment. Additionally one player held an unlabeled commodity. The other player could offer to buy this commodity. It was known that the commodity either had little value or was very precious. All gains from trade would lie with the buyer. If the seller accepted, the deal was struck. Otherwise this round ended and both players kept their endowments. If the commodity traded, Nature determined the value of the commodity. In the final stage, both players could impose harm on their counterpart, at a price to themselves. In the treatment, three stages were added. In the first new stage, after Nature had decided, a third player decided about “the appropriate purchase price”. Her decision was kept confidential. In the next two new stages, the original players had a chance to renegotiate, using the same protocol as before. If negotiations failed, the third party’s decision became effective.

We hypothesized that the rule would affect how the parties judge the fairness of the deal, both ex ante and ex post, and that this would affect their choices. Results support this prediction. If the rule is in place, buyers offer lower prices, and more deals are struck. This is why the rule turns out efficient. In the German legal discourse, the rule is mainly justified by its purported effect on ex post fairness, though. We qualify this expectation. Third parties indeed almost exclusively split gains from trade equally, despite the fact that the buyer carried all the risk and thereby insured the seller. Yet sellers themselves did not see unfairness. Hardly any seller used the punishment option. Unpredicted by the German legal debate, buyers were much more likely to exhibit ex post discontent. Their willingness to punish the seller was, however, reduced by the rule.

f) Alternative Empirical Methods

My presentation of empirical methods for behavioral legal analysis has mirrored the methods that have actually been used in the publications in *JELS*. One prominent methodological alternative has not yet made it into this journal, the field experiment. In a field experiment, the experimenter directly intervenes into the social phenomenon she wants to understand. This method is promising in that it combines high external with high internal validity. There are two main challenges. The first challenge is practical. It is technically often not easy, politically often problematic, and maybe just plain illegal if some individuals are randomly deprived of treatment the researcher herself expects to be efficient and beneficial. The second challenge is methodological. However hard the researcher tries, precisely because intervention is in the field, there is less experimental control. The validity of randomization hinges on the definition of the pool from which participants are drawn. These participants cannot be completely standardized, so that treatment effects may result from alternative causes for which the researcher must try to control. Finally, it is standard in field experiments to not reveal the manipulation. This may lead to ethical concerns.

The following is an illustration how this method can be applied to a legal issue (Listokin 2010). Used iPods were auctioned off on eBay with randomly varied return policies. Some iPods came with a satisfaction guaranteed policy, others with an explicit warranty that resembled the default warranty of the Uniform Commercial Code, and still other iPods were sold “as is.” Finally, a batch of iPods was silent regarding the return policy. The mean price paid in the auction shows that consumers are sensitive to the warranty. Prices were highest if buyers were free to give the iPod back. Prices were lowest if a guarantee was expressly excluded. In the two remaining conditions, prices were in the middle, and not statistically distinguishable. The author concludes that buyers assume a safety level as guaranteed by the majoritarian default of the code if the contract is silent on the warranty.

Economists sometimes use simulation to show on which sets of parameters a problem is conditioned. Simulation presupposes the complete definition of a mechanism. In each simulation run, one parameter is changed. Simulation is also useful if one expects one or more processes to be random, with a defined nature of the disturbance. Recently, simulation has also been used in the behavioral legal literature. In the very differentiated empirical literature on lineups, it is held to be established that eyewitnesses are more reliable if they base their recognition judgment on an absolute, rather than a relative criterion. It is argued that eyewitnesses might accept the relatively closest analogue to their recollection, rather than the one person that is so similar to memory traces that there is no reason to doubt. The paper translates this claim into a formal model of recognition and manipulates memory accuracy, the degree of similarity between perpetrator and foil, and the decision criterion. It turns out that trying to meet an absolute threshold does not minimize false positives under all circumstances. A relative approach performs better if the witness’s memory is relatively accurate and an innocent suspect is fairly similar to the perpetrator (Clark et al. 2011).

All of the foregoing methods are quantitative. They are meant to generate evidence the relevance of which is judged by way of frequentist statistics. Occasionally, law and psychology researchers instead use qualitative methods. One recent study interviewed defendants who just had pled guilty, and explored in which ways they had understood the plea inquiry. It turned out that errors were widespread. Two thirds of the sample were correct on less than 60% of the questions the researchers asked them (Redlich and Summers 2011). The main advantage of qualitative empirical research results from the fact that it is not bound by a set of dependent variables that have been defined *ex ante*. The researcher can give each individual observation full justice, and can use the evidence to learn about hitherto neglected aspects of the issue at hand. The main drawback directly results from this advantage. Since the dependent variable is not standardized, it is more difficult to assess whether individual observations generalize.

3. Future Directions

Outside criminology and law and psychology, the empirical legal movement is still young. This also holds for its behavioral branch. In conclusion I sketch potential paths for future development.

When presenting the most prominent empirical methods in greater detail, it has become clear that there is a tradeoff between external and internal validity. A straightforward reaction is combining more than one empirical method on the same research question. Occasionally, this is even done within the same publication. A study on racial bias in bankruptcy uses this approach (Braucher et al. 2012). In the US, consumers can file bankruptcy under chapter 7 or under chapter 13. While the latter procedure may be advantageous if the debtor wants to protect valuable assets, consumers usually use the former procedure since it is less onerous and less costly. The study has two parts. In the first part it shows that African Americans are disproportionately more likely to file under chapter 13, even after controlling for relevant socio-demographic factors. A vignette study randomly asks bankruptcy lawyers who usually represent consumers to give advice to a couple with Christian names that suggest an Afro-American or a Caucasian background. Bankruptcy lawyers advise the former couple significantly more frequently to file under chapter 13.

Quantitative studies quantify the trust one may have in the result by a significance test. Nonetheless, researchers may have overlooked a qualifying factor. Despite their attempts at securing randomness, experimentalists may have worked with an atypical sample. Inadvertently, a feature of the design of an experiment that seemed innocent may have been critical. Ultimately, the law should therefore be hesitant to derive normative conclusions from a single empirical study. A procedure that is standard in medicine is still very rare in law, the replication of findings (for an exception Hall 2010). Equally rare is the reanalysis of empirical data with alternative statistical models (for an exception see Goodsell et al. 2010). Finally meta-analysis, i.e. the structured, quantitative analysis of findings from a whole line of research is thus far confined to law and psychology, and almost exclusively to forensic psychology (e.g.

Stebly et al. 2011). As the field matures, all of these methods for assessing the robustness of findings and for better understanding framework conditions should become more prominent.

Economic experiments have been invented as tests for formal economic theory. Decades ago, formal economic theory has made headway into law. In many subfields of law, and in private law in particular, economic theorizing is fairly advanced. Thus far, tests of formal law and economics hypotheses are still rare. One example is a paper that first models the effect of split-award statutes on negotiations between tortfeasors and victims (Landeo et al. 2007). Under these statutes, plaintiff only receives a portion of punitive damages, while the remainder is paid to the state. The authors translate the situation into a sequential game, solve for equilibrium, and test the resulting predictions in the lab. They find that these statutes do not affect the level of care, but reduce the likelihood of trial, and the total litigation cost born by the community of parties.

Disciplines have their traditions. Traditions may result from historical contingency. Yet at least in the long run, traditions are likely to converge to the functional needs of a discipline. Arguably the different experimental traditions in psychology and economics reflect differences in the dominant research questions. The same claim seems plausible for the different approaches to the analysis of field data in criminology and econometrics. These observations suggest that, in the long run, behavioral legal researchers might want to develop their own, discipline specific empirical methods. Two challenges are likely to be particularly pronounced if the evidence shall be introduced into legal argument.

Lawyers frequently want to judge proposals for institutional intervention. This is obvious if a legal scholar makes a contribution to the legal policy discourse. Similar questions are asked by doctrinal lawyers if they have to decide on the constitutionality of institutional reform, or if they have to rely on teleological interpretation to resolve an ambiguity of the text. Often the purported effect of the intervention hinges on assumptions about the behavior of typical addressees. Yet the policy question is only partly answered by a list of relevant findings from basic behavioral research. It ultimately matters whether the specific intervention delivers on its promises, without having too many undesirable side-effects. Answering this question might require testing entire institutions, rather than isolated effects. Of course, if one does, one partly loses control since institutions are lumpy responses to lumpy perceived problems. Nonetheless, the knowledge generated that way may be more valuable since this very combination of effects is likely to be at work in legal practice. Related to this, legal institutions are very rarely designed from scratch. The typical situation is institutional reform. The legislator intervenes in the interest of improving what it thinks fell short of normative expectations. Therefore often the critical behavioral question is how addressees will react to an intervention meant to induce behavioral change. In an experiment this can be reflected by a sequential design that focuses on the difference before and after the introduction of the tested legal institution.

Ultimately legal rules are meant to decide disputes. Negotiators, administrators and courts must settle disputes in good time. To do this effectively, they must reflect what the parties see as essential features of the case. Both the characteristic time pressure and the typical level of specificity do not easily go in common with empirical methods that have been developed in the social sciences to answer questions of basic science. Non-behavioral areas of law have found solutions for the resulting methodological challenges. Merger simulations provide a good illustration (Budzinski and Ruhmer 2010). They rely on formal economic theory of industrial organization. They capitalize on rigorous econometric work in this subdiscipline of economics. But the actual simulation does not try to build all the methodological safeguards into the simulation model. And, since no closed form solution is required, simulation may simultaneously address many dimensions of a merger case, even if there is debate how to model, or how to measure them. For appropriate legal applications, behavioral researchers might develop similar tools.

In many respects, empirical behavioral research on legal issues still is a nascent endeavor. Inevitably, this chapter has only been able to provide a snapshot on a rapidly moving field. There are two take-home messages though. The field is highly differentiated and capitalizes on multiple methods from many neighboring disciplines. There is considerable room for improvement. Behavioral legal researchers should take methodological standards that have developed in the more mature neighboring fields more seriously. And they should spend more energy on developing empirical methods that directly map the needs of their own discipline.

References

- BLUNDELL, RICHARD and MONICA COSTA DIAS (2009). "Alternative Approaches to Evaluation in Empirical Microeconomics." Journal of Human Resources 44(3): 565-640.
- BRAUCHER, JEAN, DOV COHEN and ROBERT M. LAWLESS (2012). "Race, Attorney Influence, and Bankruptcy Chapter Choice." Journal of Empirical Legal Studies 9(3): 393-429.
- BRINIG, MARGARET F. and NICOLE STELLE GARNETT (2012). "Catholic Schools and Broken Windows." Journal of Empirical Legal Studies 9(2): 347-367.
- BUCCAFUSCO, CHRISTOPHER and CHRISTOPHER SPRIGMAN (2011). "Valuing Intellectual Property: An Experiment." Cornell Law Review 96(1-46).
- BUDZINSKI, OLIVER and ISABEL RUHMER (2010). "Merger Simulation in Competition Policy. A Survey." Journal of Competition Law and Economics 6(2): 277-319.
- CARDI, W. JONATHAN, RANDALL D. PENFIELD and ALBERT H. YOON (2012). "Does Tort Law Deter Individuals? A Behavioral Science Study." Journal of Empirical Legal Studies 9: ***.
- CLARK, STEVEN E., MICHAEL A. ERICKSON and JESSE BRENNEMAN (2011). "Probative Value of Absolute and Relative Judgements in Eyewitness Identification." Law and Human Behavior 35(5): 364-380.
- EISENBERG, THEODORE and CHRISTOPH ENGEL (2013). Assuring Civil Damages Adequately Deter. A Public Good Experiment
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2041154.
- ENGEL, CHRISTOPH (2005). Generating Predictability. Institutional Analysis and Institutional Design. Cambridge, Cambridge University Press.
- ENGEL, CHRISTOPH (2010). "The Behaviour of Corporate Actors. A Survey of the Empirical Literature." Journal of Institutional Economics 6: 445-475.
- ENGEL, CHRISTOPH and MICHAEL KURSCHILGEN (2011). "Fairness Ex Ante and Ex Post. Experimentally Testing Ex Post Judicial Intervention into Blockbuster Deals." Journal of Empirical Legal Studies 8: 682-708.
- FRAKES, MICHAEL (2012). "Defensive Medicine and Obstetric Practices." Journal of Empirical Legal Studies 9(3): 457-481.
- GAZAL-AYAL, OREN and RAANAN SULITZEANU-KENAN (2010). "Let My People Go: Ethnic In-Group Bias in Judicial Decisions. Evidence from a Randomized Natural Experiment." Journal of Empirical Legal Studies 7(3): 403-428.

- GOODSELL, CHARLES A., SCOTT D. GRONLUND and CURT A. CARLSON (2010). "Exploring the Sequential Lineup Advantage Using Witness." Law and Human Behavior **34**(6): 445-459.
- HALL, MATTHEW (2010). "Randomness Reconsidered. Modeling Random Judicial Assignment in the US Courts of Appeals." Journal of Empirical Legal Studies **7**(3): 574-589.
- HERTWIG, RALPH and ANDREAS ORTMANN (2001). "Experimental Practices in Economics. A Methodological Challenge for Psychologists?" Behavioral and Brain Sciences **24**(03): 383-403.
- KAHNEMAN, DANIEL and AMOS TVERSKY (2000). Choices, Values, and Frames. Cambridge, Russell Sage Foundation; Cambridge University Press.
- KLICK, JONATHAN and THOMAS STRATMANN (2007). "Diabetes Treatments and Moral Hazard." Journal of Law and Economics **50**(3): 519-538.
- LANDEO, CLAUDIA M., MAXIM NIKITIN and LINDA BABCOCK (2007). "Split Awards and Disputes. An Experimental Study of a Strategic Model of Litigation." Journal of Economic Behavior & Organization **63**(3): 553-572.
- LEAMER, EDWARD E. (1983). "Let's Take the Con out of Econometrics." American Economic Review **23**: 31-43.
- LISTOKIN, YAIR (2010). "The Meaning of Contractual Silence. A Field Experiment." Journal of Legal Analysis **2**(2): 397-416.
- POSNER, RICHARD A. (2011). Economic Analysis of Law. New York, Aspen Publishers.
- RACHLINSKI, JEFFREY J. (2011). "The Psychological Foundations of Behavioral Law and Economics." University of Illinois Law Review: 1676-1696.
- REDLICH, ALLISON D. and ALICIA SUMMERS (2011). "Voluntary, Knowing, and Intelligent Pleas. Understanding the Plea Inquiry." Psychology, Public Policy, and Law **18**: 626-643.
- REDLICH, ALLISON D., ALICIA SUMMERS and STEVEN HOOVER (2010). "Self-Reported False Confessions and False Guilty Pleas among Offenders with Mental Illness." Law and Human Behavior **34**(1): 79-90.
- ROSE, MARY R., SHARI SEIDMAN DIAMOND and KIMBERLY M. BAKER (2010). "Goffman on the Jury." Law and Human Behavior **34**(4): 310-323.
- SCHWARTZ, GARY (1994). "Reality in the Economics of Tort Law. Does Tort Law Really Deter?" UCLA Law Review **42**: 377-444.

- SIMON, DAN, DANIEL C. KRAWCZYK and KEITH J. HOLYOAK (2004). "Construction of Preferences by Constraint Satisfaction." Psychological Science **15**: 331-336.
- STEBLAY, NANCY K., JENNIFER E. DYSART and GARY L. WELLS (2011). "Seventy-Two Tests of the Sequential Lineup Superiority Effect. A Meta-Analysis and Policy Discussion." Psychology, Public Policy, and Law **17**(1): 99-139.
- SUNSTEIN, CASS R., Ed. (2000). Behavioral Law and Economics. Cambridge series on judgment and decision making. Cambridge England, Cambridge University Press.
- YANG, Y.TONY, DAVID M. STUDDERT, S. V. SUBRAMANIAN and MICHELLE M. MELLO (2012). "Does Tort Law Improve the Health of Newborns, or Miscarry? A Longitudinal Analysis of the Effect of Liability Pressure on Birth Outcomes." Journal of Empirical Legal Studies **9**(2): 217-245.
- ZAMIR, EYAL and BARAK MEDINA (2011). Law, Economics, and Morality. Oxford, Oxford University Press.
- ZAMIR, EYAL and ILANA RITOV (2012). "Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation." Journal of Legal Studies **41**(1): 165-207.
- ZEILER, KATHRYN and CHARLES R. PLOTT (2005). "The Willingness to Pay/Willingness to Accept Gap, the Endowment Effect, Subject Misconceptions and Experimental Procedures for Eliciting Valuations." American Economic Review **95**: 530-545.