

On Experiments in Empirical Legal Research

*Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.**

In this paper we discuss some strengths, stumbling blocks, common mistakes, and controversial issues that can be important when conducting experiments in the legal domain. To this end, we first briefly introduce the experimental method and note some of its strengths when used in legal research projects. We also briefly discuss important differences between laboratory and field experiments. We then examine some issues that we think are important for researchers with no or little experience in conducting empirical legal research. The list of issues discussed, although certainly not exhaustive, is intended to guide novice researchers who want to learn about experiments either because they are considering adopting this research method in their own research projects or because they want to be able to give an informed opinion about the method when reading about it in research publications or hearing about it in research presentations.

We discuss these issues because we assume that empirical research, including experimental research, may complement normative legal research. Thus, we do not argue that empirical research, in general, or experimental research, in particular, should replace normative “black letter law” but, rather, we think there are good reasons why solid empirical research may be conducive to a thorough legal science. Furthermore, although we focus on experiments in this contribution, we do not suggest that experiments are the most important empirical research method or the most promising method for the legal domain. We merely argue that experimental research is an important empirical method that deserves the attention of legal scholars and practitioners. In particular, we propose in this paper that experiments may provide fundamental insight into what is driving human behavior and what is going on in society. This insight can be important, partly because this insight may help move the legal domain to go beyond what people (including legal scholars) believe or do not believe to be true. In this way, findings from experimental studies can help to better understand core elements of the functioning of law.¹

We also note explicitly that we do not frown upon qualitative² methods, or on other quantitative³ or mixed-method⁴ research projects, quite the contrary, but here we concentrate on one particular research method, experiments, and discuss some strengths and weaknesses of this method when used in the legal domain.

* Prof. dr. Kees van den Bos, Department of Psychology and School of law, Utrecht University. Mr. Liesbeth Hulst, B.Sc., M.Sc., Department of Dutch Private Law, VU University Amsterdam.

1 More on this issue in Van den Bos 2014.

2 See, for example, Van den Bos, Loseman, and Doosje 2009.

3 See, for example, Van den Bos, Van der Velden, and Lind 2014.

4 See, for example, Hulst, and Akkermans 2011.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

Let us turn to these strengths and potential stumbling blocks of experiments in empirical legal studies.

Strengths of the Experimental Method

Empirical research tries to gain knowledge by observing what is going on in reality. This observation process can be very difficult and as a researcher you can make many errors in this process. You want to increase what you can explain with your research findings (often called “systematic variance”), and you want to decrease what you cannot explain (or “error variance”).⁵ Thus, empirical research can be depicted as a fight against error variance. An important strength of the experimental method is that it is designed to reduce error variance in important ways. The experimental method does this by formulating testable hypotheses, determining independent variables (experimental conditions)⁶ and how they are operationalized, specifying the participants that are involved and the procedure used to assign conditions to the participants, and determining the measurement of dependent variables.⁷ Box 1 gives an example of an experiment in which there is one independent variable and one dependent variable.

Many relationships of interest to empirical researchers are of the type where one variable might have an effect on another variable. For example, one might be interested in how an offender’s apology affects a number of plaintiffs’ perceptions that may influence negotiation outcomes.⁸ Or you may want to know how a seemingly trivial issue such as “what the judge ate for breakfast” has an influence on judicial decisions.⁹

5 For an explanation of these and other technical terms, see, for example, Hays 1981 and Kirk 1995.

6 Independent variables are variables that serve the function in research designs of showing whether they affect other variables, the dependent variables. Independent variables may constitute concepts that are already there, for example, variables such as age or gender that are measured among research participants. In experiments, independent variables are usually manipulated, for example in such a way that one group of research participants respond to one set of stimulus materials, whereas another group of participants respond to another set of stimulus materials. When randomly assigned to one of the two groups, the difference in reactions between the two groups of participants can be attributed to the difference in materials presented to them. The dependent variables in an experiment thus measure changes in the participants’ reactions as a result of differences in the independent variables (in this example, experimentally manipulated differences in stimuli presented to participants).

7 Kirk 1995.

8 Robbennolt, 2006.

9 See Danziger, Levav, and Avnaim-Pesso 2011. This research focuses on the debate between legal formalism and the legal realist movement. Legal formalism holds that the outcome of legal cases depends solely on laws and facts because judges apply legal reasons to the facts of a case in a rational, mechanical, and deliberative manner. In contrast, legal realists argue that the rational application of legal reasons does not sufficiently explain judicial decisions and that psychological, political, and social factors influence rulings as well. The realist view is com-

It is important to note that there may be practical or ethical barriers to using experimental methods to study some relationships of interest, for example because knowing the results of research on this issue may be considered to have unethical effects on litigants to experimentally vary whether judges recently ate or not. We will briefly touch upon this topic in the next paragraph. Here we will first continue to discuss the potential strengths of the experimental research method.

Box 1. Example of an experiment with one independent variable and one dependent variable

Recently, empirical legal researchers are getting interested in the issue of perceived procedural justice. Suppose you are interested in people's perceptions of how fairly and justly they have been treated by a decision-making legal authority and that you want to ascertain how these perceptions of perceived procedural justice affect how satisfied people are with the outcome decision they subsequently receive from the decision-making authority. In an experiment, your independent variable could involve (some aspect of) procedural justice, and your dependent variable could be participants' outcome satisfaction.¹⁰ A way to directly manipulate an important aspect of procedural justice is to vary that participants are either given the opportunity to voice their opinion about the decision that has to be made or are denied such a voice opportunity. Being allowed voice is a central component of perceived procedural justice, so this manipulation varies whether participants experience an important component of fair or unfair procedures. The experiment can then measure whether the experience of receiving voice or no-voice procedures affects participants' outcome satisfaction ratings. Thus, unlike in correlational research where one would look at how procedural justice and outcome satisfaction covary, in experimental research we manipulate one variable (independent variable) to observe its effects on another (dependent variable). In this example, the effect of being allowed or denied voice is a manipulation of procedural justice, perceptions of procedural justice serve as a check to see whether the manipulation worked, and satisfaction with the outcome that people receive is the main dependent variable.

An important strength of the experimental method is that it allows you to precisely test a causal relationship between independent and dependent variables. In its basic form, the experimental method entails independent and dependent variables, and the causal order is such that independent variables come earlier in time than do dependent variables. Thus, with experimental research, one can study the causal effect of one or more independent variables on dependent variables.

monly caricaturized by the trope that justice is "what the judge ate for breakfast." Remarkable empirical findings presented by Danziger et al. suggest that whether or not the judge recently had breakfast or another meal indeed has an effect on the judge's decisions. That is, the likelihood of a favorable ruling is greater at the very beginning of the work day or after a food break than later in the sequence of cases.

10 See Van den Bos, Lind, Vermunt, and Wilke 1997.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

Another great advantage of an experiment is the ability to ensure that the stimuli in the experimental conditions are the same and that other variables that are not part of the design (so-called “nuisance or confounding variables”) do not affect the independent and dependent variables of the experimental study. This, and the careful and controlled operationalization of the independent and dependent variables, enhances the likelihood that the results obtained in the experiment can be interpreted in a confident manner and in meaningful ways. Thus, the comparability of conditions and the controlling of nuisance are among the important strengths of the experimental method. The fact that the particular experimental procedures followed and the results obtained using these experimental procedures can be communicated in transparent ways is also a possible strength of the method.

A crucial aspect of the experimental design is that, in its ideal form, the assignment of conditions to participants is done in a random way.¹¹ When random assignment has taken place and the sample size is sufficiently large, researchers can be relatively certain that nuisance variables, such as differences in the personalities or backgrounds of the participants, are distributed evenly across conditions. Any differences between conditions thus observed on the dependent variables are likely to be due to the independent variable(s) encountered in the experiment.¹²

One of the questions in the upcoming field of empirical legal research asks what the conditions are under which perceived procedural justice is important for people. Experimental studies indicated that when people have been reminded about their personal uncertainties¹³ or when they are uncertain about how to interpret their outcomes,¹⁴ the experience of fair and just procedures (as opposed to unfair and unjust procedures) has strong effects on people’s subsequent reactions. These effects are probably there because uncertain people are inhibited how to respond, as is the case, for example, when litigants are being ordered to appear at insolvency court hearings.¹⁵ Thus, an important potential strength of the experimental design is that it can answer causal questions with precision and thereby reveal important insights into issues that might otherwise remain unnoticed or might be difficult to discover (such as the observation that conditions of uncertainty impact why perceived procedural justice can have strong effects on people’s reactions such as their trust in judges).

11 In quasi-experimental designs, non-random assignment may sometimes take place. Statistical procedures can try to correct for this, but in essence, statistical control cannot make up for flaws in methods, such as non-random assignment.

12 Wilson, Aronson, and Carlsmith 2010.

13 Van den Bos 2001a.

14 Van den Bos et al. 1997.

15 Hulst, Van den Bos, Akkermans, and Lind 2014.

Laboratory and Field Experiments

Here we do not provide a complete introduction of the experimental method¹⁶ nor a full briefing on the ins and outs of how to set up and conduct experiments.¹⁷ We also do not aim to discuss all the different types of experiments that are out there.¹⁸ This noted, we do distinguish between laboratory experiments (which are quite often experiments conducted in the psychology laboratory¹⁹) and field experiments (which are experiments conducted in more naturalistic field settings outside laboratories²⁰). While the different labels merely seem to suggest a difference in the context where the experiments are conducted, it is important to realize that the goals that researchers have with these experiments tend to be very different. That is, the primary goal of laboratory experiments, generally, is to test scientific theories, whereas the goal of field experiments tends to be enhanced methodological rigor and control in the empirical research project.

Of course, methodological rigor is also an important issue in laboratory experiments, and many field experiments are conceptually grounded in important ways, yet the primary goals researchers have with these types of experiments tend to be different. For example, a scientific researcher may be interested in why experienced procedural justice tends to outweigh outcome concerns in many survey studies conducted in the legal domain,²¹ while outcome concerns are also known to be driving people's reactions in at least some important research studies.²² The scientist may then propose that this effect occurs because in many circumstances, people are missing information about the outcomes that comparable other people receive and that people in those circumstances rely on the information that is available to them. Because how fairly and justly one has been treated is information that is quite often available, this analysis suggests that information about perceived procedural justice tends to be used as a substitute for the missing information about the outcomes of others. Thus, this conceptual analysis explains why perceived procedural justice often has strong and reliable associations with trust in judges and trust in law.²³ Box 2 describes an experiment that tests this line of reasoning using two independent variables, one dependent variable, and manipulation checks.

16 For a thorough introduction, see, for example, Kirk 1995.

17 See, for example, Smith 2000 and Wilson et al. 2010.

18 See, for example, Cook and Campbell 1979.

19 For more information about laboratory experiments, see Wilson et al. 2010.

20 For more information about field experiments, see Reis and Gosling 2010.

21 Tyler 1990.

22 Adams 1965.

23 Tyler and Huo 2002.

Box 2. Example of an experiment with two independent variables, one dependent variable, and manipulation checks

A straightforward way to test the above-mentioned conceptual analysis would be to bring research participants into situations in which they receive a certain outcome. In such a laboratory experiment, the researcher can vary in the first independent variable whether participants are either given an opportunity to voice their opinion to a decision-making authority about an outcome decision that the authority has to make or are denied such a voice opportunity. Manipulation checks can be included in the experiment to assess whether participants indeed experience the resulting voice procedure as fair and perceive the no-voice procedure as unfair. Furthermore, the manipulation of receiving or not receiving voice can be crossed with the second independent variable that varies whether participants either are or are not informed about the outcome decisions that comparable other participants receive. Dependent variables can assess how satisfied participants are with their outcome decisions. A central hypothesis that can then be tested in the lab experiment is whether the manipulation of procedural justice (i.e., voice or no-voice procedures) significantly affects participants' satisfaction ratings when participants do not know the outcome of other participants, and that the manipulation of procedural justice does *not* have an effect on satisfaction ratings when participants do know the outcome of other participants. Laboratory experiments indeed have been done in such a manner, for example those in which participants experienced receiving a certain outcome decision, were being given voice or no voice, and knew or did not know about the outcomes of comparable other persons. These lab experiments have found supportive evidence for the line of reasoning briefly described here.²⁴ These findings support a theory about why and when procedural justice and outcome concerns impact people's reactions.²⁵

We note explicitly that many of our examples thus far are about experiments that investigate the impact of perceived procedural fairness on people's reactions (see, for example, Boxes 1 and 2). These topics are important examples of the application of social psychology in a legal context, and results of such studies can be taken into account by legal practitioners. However, other topics studied by means of experimental manipulations other than what we focus on here are also important and should be taken into account when examining the whole spectrum of behaviors and reactions in the legal domain. To this end, Boxes 3 and 4 describe examples of experimental studies from other domains of law. Box 3 gives an example of a laboratory experiment on how people interpret contractual obligations, and Box 4 describes an example of a field experiment in criminal law about the effects of different photo lineups among actual eyewitnesses.

²⁴ For details, see Van den Bos et al. 1997.

²⁵ Van den Bos and Lind 2002.

Box 3. Example of a laboratory experiment on contract law

This example focuses on how people design and evaluate a construction contract. For instance, imagine that parties have the option to design a construction contract with a price of 80 euros and a bonus of 20 euros for timely performance, or with a price of 100 euros and a penalty of 20 euros for late performance. From a rational choice perspective, the two contracts are identical. However, framing theories in psychology suggest that a penalty of 20 euros will be perceived as more aversive than the foregone chance to earn a bonus of 20 euros. Hence, framing theories would predict a higher incidence of timely performance under penalty contracts. Framing may also influence the way people interpret ambiguous contractual obligations. This is important because contracts often leave room for parties to decide what precisely their contractual duties are. Recent laboratory experiments tested the prediction that people will tend to adopt a more selfish interpretation of their contractual obligations when they are trying to minimize losses as opposed to enhancing profits.²⁶ Thus, the idea tested in these lab experiments was that in situations in which people can interpret their duties in more than one way, loss frames (installed, for example, by means of penalty contracts) would make parties adopt an interpretation that is more self-serving. To test this hypothesis, several experiments were conducted. One laboratory experiment among students of Hebrew University of Jerusalem included two conditions. Participants in the gains condition of this experiment were told that they would be asked to answer 20 trivia questions and that for each correct answer they would receive 1 NIS. Participants in the loss condition were informed that they would receive 20 NIS for their participation in the experiment, but for each mistake in the trivia quiz they would lose 1 NIS. Results obtained from this lab experiment indicate that participants in the Gains condition chose to solve more difficult questions than those in the Losses condition. According to the authors of this research project, this effect suggests that under conditions of loss framing, people are more inclined to adopt a more selfish perspective to minimize their losses than under conditions of gain framing. The authors consider the findings obtained in this laboratory experiment to be supportive of their idea that framing contractual payoffs as losses rather than as gains raises parties' tendency to interpret their obligations selfishly.²⁷

26 See the third study presented in Feldman, Schurr, and Teichman 2013.

27 For details, see Feldman et al. 2013.

Box 4. Example of a field experiment in criminal law

An important issue in criminal law is how to present photos of possible perpetrators to eyewitnesses. For instance, a sequential procedure can be used in which the witness views lineup members one at a time and makes a decision on each before seeing the next. This contrasts with a simultaneous procedure in which all lineup members are available to be viewed at the same time. In a field experiment among actual eyewitnesses to actual crimes in 4 police jurisdictions in the United States were randomly assigned to view simultaneous or sequential photo lineups using laptop computers and double-blind administration.²⁸ The findings obtained in this experiment yielded no statistically significant effects on rates of identifying lineup suspects, but the sequential procedure produced a significantly lower rate of identifying known-innocent lineup fillers than did the simultaneous procedure. According to the authors of this research project, these results suggest that the sequential procedure that is used in the field reduces the identification of stimulus persons known to be innocent, but the authors also note that the differences observed are relatively small.²⁹

Now that we have given some examples of laboratory and field experiments, it is important to emphasize that theory construction and testing are the main goals of laboratory experiments. The artificial stimulus materials used to test the theory limit what we learn from the findings reported of such lab experiments. That said, the theory thus tested can be used to generalize and speculate about the relevance of the theory for what is going on in circumstances other than the specific context in which it has been tested.³⁰ Importantly, given the enormous variety of contexts that are present in the legal world and that many different situations can be devised to test theories, it is nearly always impossible to test each and every circumstance. Having a solid theory that has survived crucial empirical tests in controlled laboratory experiments can thus be a pivotal tool to form solid and basic insight into what can be assumed to be going on in the domains of law and society and human behavior. That is why researchers who use lab experiments stress that experiments should be vivid and “psychologically real” to research participants and that mundane realism is less important for their goals.³¹ Many lab experimenters may overdo this,³² but the thing is that laboratory experiments and the theory construction that is associated with these experiments can and should serve to better ground the domains of law and human behavior and society. Thus, while in the remainder of this text we focus mainly on field experiments in the

28 Wells, Steblay, and Dysart 2015.

29 For details, please see Wells et al. 2015.

30 Aronson, Wilson, and Akert 2013.

31 Wilson et al. 2010.

32 For example, lab experimenters may neglect mundane realism too much and may think that their experiments are psychologically very real and more vivid to participants than they in reality are. The discussion of these and other issues could easily be a subject for an entirely different paper that well extends the space allotted to the current paper. See also Ring 1967 and Gergen 1973, 1978 as well as McGuire 1967 and Wallach and Wallach 1994.

legal domain, we do note the relevance of more basic kinds of experiments conducted in controlled settings such as the psychology laboratory.

Of course, it may well be possible that an empirical legal researcher would prefer to conduct a field experiment to test the above-mentioned line of reasoning in important real-life contexts, such as the courtroom. We certainly understand this and indeed we would applaud a greater usage of field experiments in both empirical legal studies and psychology. In fact, field experiments may be the ultimate and most advanced research method available to empirical legal researchers. After all, when done properly, field experiments can test sophisticated scientific theories by means of rigorous yet ethically sound experimental manipulations in meaningful field settings such as courtroom hearings.³³ These kinds of experiments can make it possible to be able to turn the knobs of what drives human behavior and the potential of revealing more precise insight into what is happening in the courtroom and in other legal domains.

This noted, it can be very hard or unethical to experimentally vary the conditions in real-life contexts such as courtroom hearings.³⁴ For example, varying whether a litigant has complete access to information about the outcome of comparable other cases and fully understands this jurisprudence may be hard to achieve and may involve years of training (comparable to the training that lawyers follow). And varying whether litigants are treated in unfair (as opposed to fair) ways by judges in the courtroom can well be viewed as inappropriate and ethically wrong. Basically, there are two solutions to these kinds of important practical and ethical issues. The first solution would be to decide to keep focusing on the problem and field setting under consideration and accept the usage of a research method that involves less methodological rigor and control. For example, rather than experimentally testing how “what the judge ate for breakfast” affects their judicial decisions, researchers might choose to study how the ordinal position of the court case is associated with the favorability of the judge’s decision and the judge having taken a break to eat.³⁵ A second solution would be to accept a somewhat more artificial or hypothetical quality of stimulus materials and to decide to use an experimental method that affords high levels of methodological control. The first solution focuses on “external validity”, and the second solution focuses on “internal validity”.³⁶

In essence, building a successful and impactful program of research involves finding the right balance between appropriate levels of “internal” and “external validity” as well as the correct focus on both problem-oriented (“bottom-up”) and theory-oriented (“top-down”) research projects.³⁷ We argue here that the field of psychological science is off balance because it focuses too much on theory-oriented projects with relatively low levels of external validity and low societal impact, whereas the imbalance in the field of empirical legal studies is such that

33 Hulst et al. 2014.

34 For an exception, see Hulst et al. 2014.

35 Danziger et al. 2011.

36 More on issues of validity in Brewer 2000.

37 See, for example, West, Biesanz, and Pitts 2000.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

there is too much emphasis on problem-focused projects with relatively low levels of internal validity and absence of causal control. It is time that scientific domains get more balanced. We hope that a conceptually grounded and methodologically thorough integration of law and social psychological experimentation may help in this process.

Focus, Focus, Focus

The legal practice is multi-faceted and complex, and the construction and interpretation of laws as well as the deep thinking demanded by legal issues involve some of the most complicated matters known to scientific scholars. Indeed, this quality is one of the core aspects of what makes the study of law so fascinating. This quality also tends to provide an extra challenge for those researchers who want to study empirical aspects revolving around the issues of law and society and human behavior. That is, precisely because legal problems tend to involve so many issues, it can be hard in empirical legal studies to focus on the core aspect of the problem under consideration. When setting up experiments, it is important to realize this because a good, successful experiment starts with identifying the “simple effect” on which the experiment is build.

A simple effect is the effect of an independent variable at a single level of another independent variable. In the example we discussed in Box 2, the simple effect was obtained by comparing how research participants responded to being allowed versus being denied an opportunity to voice their opinions.³⁸ Varying and comparing these two conditions of procedural justice thus constituted the building block on which the experiment was based. The other manipulated independent variable, which varied whether participants did or did not know about the outcome of comparable other persons, served as a moderator variable, and the hypothesis in the experiment examined whether participants’ reactions to the voice versus no-voice manipulation were different as a function of whether participants knew or did not know the outcome of comparable other persons. In this way, the moderator variable qualifies the effects of the procedural justice manipulation.

In laboratory experiments, the simple effect is generally the building block of the experimental design to which other independent variables (“moderators”) are added. Related to this, in field experiments, the simple effect should give direction to researchers regarding what to focus on in the field context in which so many things may be going on or might be important. Thus, our advice is to focus in both laboratory and field experiments on variables that really matter and that are really important for your line of reasoning. Involve in your design some independent variables, dependent variables, and attempt to control for some nuisance or confounding variables, and focus on those variables. Remember that no matter how difficult this may be, a good, successful empirical research project tends to

38 Van den Bos et al. 1997.

focus on core issues under consideration that are conceptually important and that can be studied and operationalized in empirically meaningful ways and to leave it at that.

One has to accept that one empirical study cannot address everything. Thus, one experiment, or even a series of experiments (such as a multi-study paper or a dissertation that includes several multi-experiment chapters), has some important limitations and cannot solve all the problems related to those limitations. The general discussion of the report one writes about an empirical research project is the place to acknowledge and accept these limitations explicitly. Doing so in a straightforward fashion will strengthen, not weaken, one's line of reasoning. We realize that focusing in this way on some variables only can be very difficult for empirical legal researchers, in part because legal training involves learning to concentrate on differences between cases and to pay special attention to the unique aspects of a particular case under consideration. This individualistic kind of case orientation and the strong attention to the many issues associated with legal problems can make it very hard to focus only on the core aspects under consideration. Nevertheless, focusing successfully is a pivotal aspect of a good empirical research project, we argue, and the importance of this cannot be overestimated.

Be Constructive

Probably because of its emphasis on detail and its attention to individual cases, we observe reluctance in the field of law to generalize and draw abstract conclusions that generalize across different cases. Surely, on the basis of individual cases lawyers and legal scholars identify general normative principles that can be applied to other cases as well. This noted, an important aim when conducting experiments frequently entails to formulate conceptual conclusions that generalize³⁹ beyond the empirical observations of the individual experiment and that describe connections between abstract concepts. Thus, whenever possible, the goal is to propose tentative conceptual relationships that ideally hold up in at least somewhat different contexts with somewhat different research participants, different stimulus materials, and different problems and issues at hand. Empirical researchers can learn from the attention to detail and differences in individual cases from classic legal research, and legal scholars can profit from the aim of experimental researchers to articulate abstract conceptual conclusions that are generalizable across cases. We recommend that empirical researchers adopt constructive attitudes, especially when setting up experimental research. Such a constructive way of approaching things is needed in processes of generalization and to formulate theories that propose that different problems share important conceptual similarities to one other.

39 West et al. 2000.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

Related to this, when important effects have been observed in some countries or contexts yet have not been studied before in other countries or contexts there is a tendency in the legal domain to conclude that we have no insight whatsoever whether the effect observed in earlier contexts works in the new contexts as well. It is fine when one scrutinizes the literature and observes that some effects have not been examined in some contexts. However, this should be treated as a starting point for possible empirical research. Importantly, the empirical researcher should then develop a line of reasoning (ideally culminating in grounded and testable research hypotheses) that argues that the effect earlier observed is or is not there in the new contexts. We note that many legal scholars skip this phase of grounding why certain effects may or may not be there in new contexts. Thus, one argues, for example, that the fair process effect discussed earlier has been observed in many different contexts but not in this new law in this particular country in this special court case. Again, this is fine as a starting point, but should be combined with a careful line of reasoning arguing why the fair process effect is unlikely to hold in the new context. After all, relationships that have not been studied before are no indication whatsoever that these relationships do not exist. Related to this, each country's legal system tends to have unique aspects, but this does not imply that many countries do not overlap with respect to important legal matters and how citizens respond to those matters in important ways. Adopting a constructive attitude can help to circumvent making inference errors and can prevent equating unexplored effects with non-existent effects.

A constructive point of view is also needed, we argue, because it may well be argued that each scientific domain consists of many not very interesting research studies. In fact, many of them may be not very good. This so-called Sturgeon's Law is supposed to apply to each and every field of scientific investigation,⁴⁰ including the fields of law and psychology. Importantly, building on Dennett,⁴¹ we suggest that when you want to criticize a field please do not waste your time on focusing on the crap. Go after the good issues and the interesting stuff, and then try to understand these issues and explain them better than has been done before. Adopting a positive, constructive approach to experiments and using this empirical method to formulate interesting and testable research predictions that have the potential to stimulate an entire field of research may be the way to go here.

Think!

Building on the previous issues, we stress that, when planning and conducting experiments, researchers need to think on their own (as opposed to following or applying insights developed by others). It is especially important to think independently about the design of your study and, particularly, how your research

40 Retrieved from http://www.openculture.com/2013/05/philosopher_daniel_dennett_presents_seven_tools_for_critical_thinking.html on August 3, 2015.

41 Dennett 2013.

participants will react toward the stimulus materials of the experiment. In our experience, this latter aspect is quite often underestimated. Careful pilot testing of the stimulus materials among a subsample of your population may be important here. However, thinking about it yourself as a researcher is even more important and is a valuable and indispensable tool toward well-developed stimulus materials that have a decent chance of being processed by your research participants in ways such that they can indeed test your conceptual predictions in ways the experiment intended. For example, researchers can try to imagine whether the wording of questions is unambiguous, also for lower-educated participants in a field experiment, and how participants will react to experimental manipulations.

Developed in this way, findings from experimental studies may help move the legal domain to go beyond what people (including legal scholars) believe or do not believe to be true. When conducted in this manner, experiments may play an important role in the field of legal research, furthering the pursuit of a basic empirical legal science. In our view, empirical legal studies not only are an applied branch of science, but can or should be a basic science as well, thus encompassing scientific endeavors oriented toward developing new, fundamental insight into what is driving human behavior and what is going on in society that is important to understand core elements of the functioning of law. Thinking in independent ways when setting up experiments in the legal domain is crucial for this development.

The Problem of the Null Effect

Importantly, no empirical method is free from problems. Thus, we note here explicitly that the experimental method has some important advantages, but some key drawbacks as well.⁴² Perhaps one of the most important problems with experiments is how to interpret effects observed in an experiment that are statistically not significant. Particularly when the effects were statistically significant in other experiments, this can yield some big problems and controversial issues. Ideally, the independent and dependent variables in any experiment are operationalized in such a careful way that the experiment can be repeated in different contexts or with different participants. When earlier results are found to replicate in different contexts or with different participants, this is indicative of the robustness of the hypotheses tested. When the results do not replicate, this ideally tells us something about the meaningful differences between the old and the new contexts studied or the old versus new groups of participants studied. As such, the clear message that we would like to convey is that non-replication is not a problem, but quite often should be treated as a starting point for more nuanced and better insight into the issues at hand.

42 See, for example, Gergen 1978; Harré 1974; Hayden and Andersen 1979; McGuire 1967, 1973; Orne 1962.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

That said, an important problem with finding effects in experiments that are statistically not significant is how to interpret these “null effects.” Perhaps, something in the operationalization process went wrong that is not so much a problem for the theory that is being tested but that is more specifically related to empirical and non-conceptual issues. For instance, in experiments in which participants receive or are denied voice opportunities, some participants may not be paying enough attention to the stimulus materials to notice that they were allowed or denied something important. Perhaps presenting the materials in ways in which research participants can process more easily can remedy this problem. If this is the case, we have learned something important, namely, how stimulus materials should be presented in such a way that concepts the experiments intended to test can indeed be tested in a meaningful way. The conclusion from such a null effect would not be that the conceptual theory is fundamentally flawed, but rather that stimulus materials should be presented in particular ways.

More generally, the problem with statistically non-significant effects in experimental research is that many different issues can be responsible for these “null effects.” The main solution to this problem is to conduct more, carefully operationalized, and thought-through research. Therefore, we emphasize the importance of programmatic research and systematic replications. We further note the preference for multi-study papers (if possible) and a multi-method approach to lines of research and research programs of individual researchers and groups of researchers.

Conclusions

This paper did not aim to provide a complete account of the do’s and don’ts of experimental research, nor was it our goal to provide a full review of the experimental method in the legal domain. Rather, we wanted to convey some basic observations that we think can be of help when setting up experiments on the interface between law and society and human behavior. In doing so, we realize that some of the observations put forward in this paper can be easily applied to other research methods as well, but we also think the issues discussed here have special relevance to carrying out experiments in empirical legal studies.

We note explicitly that we focused on experiments from research in perceived procedural justice (see Boxes 1 and 2), contract law (see Box 3), and eyewitness reports in criminal law (Box 4). Although interesting and informative, a wider variety of experiments on other issues pertaining to the legal domain would be important to consider when contemplating how and why to conduct legal experiments. Consider, for example, recent experimental studies in the field of criminal law that focused on the role of familiarity in identification.⁴³ Furthermore, when considering the domain of civil law, it is noteworthy to highlight interesting

43 Searston, Tangen, and Eva 2016.

recent experimental research that used public good games to study whether different civil damage regimes affect deterrence and cooperation.⁴⁴ Other experiments related to civil law examined sexism in labor arbitration decisions⁴⁵ or used scenarios to assess the effectiveness of disclaimers in mutual fund performance advertisements.⁴⁶ Again, we emphasize the possible useful role of experimental research in various domains of law.

We think that overcoming the stumbling blocks identified in this paper may lead empirical legal researchers to profit from the strengths of the experimental method. Among the strengths of the experimental approach is enhanced methodological control in research design and research questions studied. Another possible strength may be the creativity and methodological rigor of experimental manipulations that may lead empirical researchers to come up with new, previously unidentified research topics.⁴⁷ These issues may not be discovered by means of other research methods. Thus, we hope the current observations are of some use for the further development of the field of empirical legal research. The insights thus obtained may help jurists both at the university and in legal practice.

References

Adams 1965

J. S. Adams, Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). New York: Academic Press. 1965

Aronson, Wilson, and Akert 2013

E. Aronson, T. D. Wilson, and R. M. Akert, *Social psychology* (8th international edition). Harlow, England: Pearson. 2013

Brewer 2000

M. B. Brewer, Research design and issues of validity. In H. T. Reis and C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 3-16). Cambridge, UK: Cambridge University Press. 2000

Cook and Campbell 1979

T. D. Cook and D. T. Campbell, *Quasi-experimentation: Design and analysis for field settings*. Rand McNally: Chicago. 1979

Danziger, Levav, and Avnaim-Pesso 2011

S. Danziger, J. Levav, L. Avnaim-Pesso, Extraneous factors in judicial decisions, *Proceedings of the National Academy of Sciences of the United States of America*, 17, 6889-6892. 2011

Dennett 2013

D. C. Dennett, *Intuition pumps and other tools for thinking*. New York: W. W. Norton. 2013

44 Eisenberg and Engel 2014.

45 Girvan, Deason, and Borgida 2015.

46 Mercer, Palmiter, and Taha 2010.

47 Van den Bos 2001b.

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

Eisenberg and Engel 2014

Th. Eisenberg and C. Engel, Assuring civil damages adequately deter: A public good experiment, *Journal of Empirical Legal Studies*, 11, 301-349. 2014

Feldman, Schurr, and Teichman 2013

Y. Feldman, A. Schurr, and D. Teichman, Reference points and contractual choices: An experimental examination, *Journal of Empirical Legal Studies*, 10, 512-541. 2013

Gergen 1973

K. J. Gergen, Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309-320. 1973

Gergen 1978

K. J. Gergen, Experimentation in social psychology: A reappraisal, *European Journal of Social Psychology*, 8, 507-527. 1978

Girvan, Deason, and Borgida 2015

E. J. Girvan, G. M. Deason, and E. Borgida, The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions, *Law and Human Behavior*, 39, 525-537. 2015.

Harré 1974

R. Harré, Some remarks on "rule" as a scientific concept, In T. Mischel (Ed.), *Understanding other persons* (pp. 143-184). Oxford: Blackwell. 1974

Hayden and Andersen 1979

R. M. Hayden and J. K. Andersen, On the evaluation of procedural systems in laboratory experiments: A critique of Thibaut and Walker, *Law and Human Behavior*, 3, 21-38. 1979

Hays 1981

W. L. Hays. *Statistics* (3rd ed.). New York: Holt-Saunders. 1981

Hulst and Akkermans 2011

L. Hulst and A. J. Akkermans, Can money symbolize acknowledgment? How victims' relatives perceive an award for their emotional harm, *Psychological Injury and the Law*, 4, 245-262. 2011

Hulst, Van den Bos, Akkermans, and Lind 2014

L. Hulst, K. van den Bos, A. Akkermans, and E. A. Lind, *Behavioral disinhibition can weaken the fair process effect on trust in judges and can uncover hidden discontent with the status quo*. Paper presented at the Fifteenth International Conference on Social Justice Research, New York, USA. 2014, June

Kirk 1995

R. E. Kirk, *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole. 1995

McGuire 1967

W. J. McGuire, Some impending reorientations in social psychology, *Journal of Experimental Social Psychology*, 3, 124-139. 1967

McGuire 1973

W. J. McGuire, The Yin and Yang of progress in social psychology: Seven koan, *Journal of Personality and Social Psychology*, 26, 446-456. 1973

Mercer, Palmiter, and Taha 2010

M. Mercer, A. R. Palmiter, and A. E. Taha, Worthless warnings? Testing the effectiveness of disclaimers in mutual fund advertisements, *Journal of Empirical Legal Studies*, 7, 429-459. 2010

Orne 1962

M. T. Orne, On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications, *American Psychologist*, 17, 776-783. 1962

Reis and Gosling 2010

H. T. Reis and S. D. Gosling, Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 82-114). Hoboken, NJ: Wiley. 2010

Ring 1967

K. Ring, K., Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, 3, 113-123. 1967

Robbennolt 2006

J. K. Robbennolt, Apologies and settlement levers. *Journal of Empirical Legal Studies*, 3, 333-373. 2006

Searston, Tangen, and Eva 2016

R. Searston, J. Tangen, and K. Eva, Putting bias into context: The role of familiarity in identification, *Law and Human Behavior*, 40, 50-64. 2016.

Smith 2000

E. R. Smith, Research design. In H. T. Reis and C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 17-39). Cambridge, UK: Cambridge University Press. 2000

Tyler 1990

T. R. Tyler, *Why do people obey the law? Procedural justice, legitimacy, and compliance*. New Haven, CT: Yale University Press. 1990

Tyler and Huo 2002

T. R. Tyler and Y. J. Huo, *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation. 2002

Van den Bos 2001a

K. van den Bos, Uncertainty management: The influence of uncertainty salience on reactions to perceived procedural fairness. *Journal of Personality and Social Psychology*, 80, 931-941. 2001

Van den Bos 2001b

K. van den Bos, Fundamental research by means of laboratory experiments is essential for a better understanding of organizational justice. *Journal of Vocational Behavior*, 58, 254-259. 2001

Van den Bos 2014

K. van den Bos, *Kijken naar het recht*. Inaugural lecture, Utrecht University. 2014

Van den Bos, Lind, Vermunt, and Wilke 1997

K. van den Bos, E. A. Lind, R. Vermunt, and H. A. M. Wilke, How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. *Journal of Personality and Social Psychology*, 72, 1034-1046. 1997

Van den Bos, Loseman, and Doosje 2009

K. van den Bos, A. Loseman, and B. Doosje, *Waarom jongeren radicaliseren en sympathie krijgen voor terrorisme: Onrechtvaardigheid, onzekerheid en bedreigde groepen*. The Hague: Research and Documentation Centre of the Dutch Ministry of Justice. 2009

Van den Bos, Van der Velden, and Lind 2014

K. van den Bos, L. van der Velden, and E. A. Lind, On the role of perceived procedural justice in citizens' reactions to government decisions and the handling of conflicts. *Utrecht Law Review*, 10(4), 1-26. 2014

Wallach and Wallach 1994

L. Wallach and M. A. Wallach, Gergen versus the mainstream: Are hypotheses in social psychology subject to empirical test? *Journal of Personality and Social Psychology*, 67, 233-242. 1994

Prof. dr. Kees van den Bos & mr. Liesbeth Hulst, B.Sc., M.Sc.*

Wells, Steblay, and Dysart 2015

G. L. Wells, N. K. Steblay, and J. E. Dysart, Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure, *Law and Human Behavior*, 39, 1-14. 2015.

West, Biesanz, and Pitts 2000

S. G. West, J. C. Biesanz, and S. C. Pitts, Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis and C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40-84). Cambridge, UK: Cambridge University Press. 2000

Wilson, Aronson, and Carlsmith 2010

T. D. Wilson, E. Aronson, and K. Carlsmith, The art of laboratory experimentation. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 51-81). Hoboken, NJ: Wiley. 2010