



**What the Judge Argues is Not  
What the Judge Thinks**

**Eye Tracking as a Window into  
Judicial Decision Making**

**Christoph Engel**

**Rima Maria Rahal**





# **What the Judge Argues is Not What the Judge Thinks Eye Tracking as a Window into Judicial Decision Making**

**Christoph Engel & Rima Maria Rahal**

January 2020

revised June 2022

# What the Judge Argues is Not What the Judge Thinks Eye Tracking as a Window into Judicial Decision Making

Christoph Engel & Rima Maria Rahal

## Abstract

Legal cases are frequently inconclusive. One source of inconclusiveness is the necessity to balance conceptually incompatible normative considerations. Still the judiciary seems to do so reasonably well. How can it? In this study, we exploit eye tracking as a window into the mental process. A first study tests whether information processing reflects the degree of normative conflict, and whether it is influenced by the ultimate disposition of the case, which is induced by assigning participants an adversarial role. These expectations are not borne out by the data. But the number and the duration of fixations on the features of the case, as recorded with eye tracking, accurately predict the disposition. A second study builds on the psychological theory of parallel constraint satisfaction. The theory posits that ambiguous problems are made tractable by gradually transforming the inputs until a coherent representation emerges. The mental process can be inferred from the explicit reevaluation of the inputs after the decision has been made. The study combines this method with eye tracking. Both measures predict outcomes with almost the same accuracy, but are uncorrelated. Eye tracking shows that the process of finding the solution and the explicit representation of the outcome are distinct mental activities.

*JEL:* D01, D81, D91, K13, K40

*Keywords:* legal decision-making, ambiguity, balancing, parallel constraint satisfaction, coherence shift, motivated reasoning, cognitive dissonance, eye tracking

## I. Introduction

Judges are no machines. They do not calculate the outcomes of legal cases, they decide. What this statement implies is best characterized in comparison with the standard approach in law and economics (see only Cooter & Ulen, 2012; R. A. Posner, 2014). In a classic model of, say, torts, a would-be tortfeasor precisely knows her own action space, and can perfectly predict the effect of choosing any action from this set on her own utility. In the simple most model, the judiciary has the same knowledge and has power to inflict a sanction if the individual chooses an action that is socially undesirable. The judge exercises her authority as a faithful agent of the legislator, choosing the sanction that induces the defendant to shift from the individually to the socially optimal choice.

The typical situation in an actual court room is very different. Often, already the facts are contested. The court must formally construct the case. In this project, we bracket this source of unpredictability, and focus on the normative side. Paying damages may serve multiple purposes (for an indepth discussion see Oberdiek, 2008, 2014; G. Schwartz, 1994; G. T. Schwartz, 1996). In first approximation, one may distinguish compensation, deterrence, and retribution (M. A. Geistfeld, 2011). From a compensation perspective, damages are correlates of property. If a good is A's property and not B's, it is for A to decide what to do with the thing. If B impinges upon A's sphere of influence, and this makes the good less valuable for A, A must be made whole (more from M. Geistfeld, 2014; Goldberg, Sebok, & Zipursky, 2016). The law may alternatively aim at prevention, rather than reparation (Eisenberg & Engel, 2014, 2016). In this forward-looking perspective, the prospect of paying damages deters rule violations (Landes & Posner, 1987). Yet arguably, people do not only care about outcomes. They also care about principles. By impinging upon another individual's property, the perpetrator has not paid the victim the respect she deserves. The law may step in as society's agent. Having the tortfeasor pay is a way of visibly expressing disdain.

These alternative normative concerns need not support the same disposition of the case. If the victim is made whole, the tortfeasor may be bankrupt: what is good for distribution is bad for incentives (as tort law becomes overly deterrent). If compensation is cheap, a wealthy tortfeasor may just do it for the fun of it: she is not deterred, and the obligation to compensate the victim does not restore the perceived justice balance. If the tortfeasor has visibly had good intentions, compensation may seem unjust, even if this leaves the damage with the victim. In terms of first principles, distributive justice, welfare, and identity are distinct. More importantly, there is no overarching normative theory that would enable the court to transparently trade one normative concern against the other (Engel, 2001). Deciding the case requires balancing (cf. Aleinikoff, 1987). There is a long-standing normative debate at the meta level about the desirability of balancing (Alexy, 2003; Petersen, 2017; Pildes, 1993). Yet, for legal decision-making in practice, there is no way around. We are interested in how decision-makers actually weigh the conflicting features of the case.

Why can legal decision making not be reduced to a mechanical act? In terms of decision theory, the typical legal case is ill-defined. This term defines a problem class. The decision

problem is only incompletely defined (Lynch, Ashley, Alevan, & Pinkwart, 2006; Newell & Simon, 1972; Reitman, 1965). The problem has more than one acceptable solution (Dörner & Funke, 2017; Kitchner, 1983). Finding the solution requires creativity (Welter, Jaarsveld, & Lachmann, 2017). Expertise matters (Schunn, McGregor, & Saner, 2005), which is how judicial experience impacts on case disposition. Most importantly for the law: the person of the judge cannot be taken out of the equation. The individual judge (or panel) cannot be fully controlled: by the appellate court, the legislator, or society at large. There is scope for bias, for instance on racial or ideological grounds. Documenting the incidence of bias, and exploring the power of solutions, is a valid enterprise. But before one can hope to make progress with debiasing, one must understand the mental mechanisms. This question of basic research is the topic of the present paper.

During the German revolution of 1848, a prominent tune claimed: “but the thoughts are free”.<sup>1</sup> No longer. With eye tracking this generation has a chance to see choices in the making, without interrupting or interfering with the decision process itself. We exploit the power of this method to gain a deeper understanding of legal decision-making, if the decision-maker is faced with an ambiguous case.

The method, widely used to study cognitive processes involved in reading (Rayner, 1998), developmental (Gredebäck, Johnson, & von Hofsten, 2009), organizational (Meißner & Oll, 2019), behavioral economics (Krajbich & Dean, 2015), as well as marketing research (Wedel, Pieters, & van der Lans, 2019), builds on the eye-mind hypothesis (Just & Carpenter, 1980), positing a relationship between eye gaze and thinking. While there is no one-to-one mapping between the two (see M. I. Posner, 1980), they are related in a probabilistic fashion (Huettig, Olivers, & Hartsuiker, 2011, p. 141): The most active location in working memory will eventually determine the most likely direction of eye movements. When factors capable of exogenously capturing visual attention (e.g., movement, differences in luminosity, contrast, position, see Orquin and Loose (2013)) are accounted for, eye movements can be used to infer voluntary shifts in attention (for overviews see Corbetta & Shulman, 2002; Theeuwes, 2010; Yantis, 2000), reflecting decision-makers’ motives, preferences, goals and abilities as they engage with a decision situation.

Measurements of eye gaze offer a versatile set of variables to gain closer insights about the cognitive processes ongoing during decision formation. Here, we chiefly rely on assessing fixation counts as an indicator of the extent to which information about the decision situation was processed, where higher numbers of fixations represent more intense processing (Just & Carpenter, 1980; Rayner, 1998). In particular, fixation counts on different pieces of information relevant to the decision situation allow inferences about the relative importance of these pieces of information for the decision maker: more important aspects are gazed at more often (Russo & Leclerc, 1994) and longer (Armell, Beaumel, & Rangel, 2008). The time spent fixating a particular piece of information (dwell time) is also used as a measure of the processing demands: more difficulty in engaging with a certain stimulus

---

1 [https://en.wikipedia.org/wiki/Die\\_Gedanken\\_sind\\_frei](https://en.wikipedia.org/wiki/Die_Gedanken_sind_frei).

is reflected in longer fixation durations (Becker, 2011; Shen, Reingold, Pomplun, & Williams, 2003).

Specifically, we use eye tracking as a window into the mental process of judicial decision-making, in two related eye tracking studies. In our first study, we test whether the number and the duration of fixations depend on the degree of normative conflict, and whether fixations focus on the features of the case that are critical for the normative conflict. We further test whether participants predominantly fixate on the items that support their preferred disposition of the case. Both predictions are not supported by the data. We additionally ask participants to evaluate the normative importance of each of the eight features of the case. Eye tracking data and these evaluations turn out to be uncorrelated.

Yet eye gaze has strong predictive power for the subsequent decision, as have the explicit ex post evaluations. This suggests that the context of discovery (observed with eye tracking) and the context of representation (observed with explicit ratings) fall apart. This can be interpreted as an empirical contribution to a debate that originated in analytic philosophy (Popper, 1935, p. 113; Ziman, 2000). Generating a decision and representing it to its respective audience(s) are independent activities (Luhmann, 1966). The normative debate about representation norms tends to consider this insight from a public choice perspective. One is concerned that biased decision-makers might hide normatively unacceptable motives behind seemingly innocent justifications (Schauer, 1995). We are not arguing that this risk should be neglected. Yet our data suggests that the distinction even matters if the decision-maker is perfectly good-natured, and exclusively motivated to implement the intentions of the law. The effect results from the fact that the actual decision-making process is not available to introspection (Engel, Timme, & Glöckner, 2020; Glöckner & Engel, 2013; Nisbett & Wilson, 1977), as cognitions about the thought process are overlaid with post-hoc rationalizations or other interferences of retrieval and reporting.

As the disconnect between eye gaze and explicit evaluations of the features of the case was unexpected, we have deemed it important to replicate the result. We have implemented our second study as a conceptual replication. The second study capitalizes on a paradigm that has originally been developed to document the constructive nature of decision-making. The theory of parallel constraint satisfaction posits that decision-makers make an ill-defined problem tractable by subconsciously reinterpreting the ambiguous evidence until a consistent story emerges.<sup>2</sup> One way of demonstrating the effect has participants evaluate the ambiguous inputs twice. During the first evaluation, inputs are presented in isolation, and without disclosing that the same outputs will later be used as descriptions of a decision problem. Thereafter participants are asked to decide. In the final step, they are requested to once more evaluate the same items. Mental process is inferred from a change in the evaluation, which this theory calls a coherence shift. In our second study, we test whether the number and the duration of fixations reflect this drift in evaluation.

We implement one additional manipulation, to test for two complementary psychological effects. For half of the participants, we induce an outcome, by assigning them to represent

---

2 The relevant literature is reported in section III.2 below.

plaintiff. The other half of participants is in the role of judges, and hence not induced to produce an exogenously defined outcome. We use the “conflict” version of the case, which is stacked against plaintiff. For participants in the role of attorneys, a coherence shift, and potentially a drift in eye gaze, could result from motivated reasoning (Kunda, 1990). This psychological concept predicts that decision-makers exploit ambiguity (which our design deliberately introduces) to bolster self-esteem. If motivated reasoning is at work, we should find reevaluations and corresponding eye gaze, but only if the participant has randomly been assigned to represent plaintiff. Alternatively, participants could engage in the reduction of cognitive dissonance (Festinger, 1957). In this perspective, they dislike ambiguity, as it casts doubt on whatever decision they make. If this mental mechanism is at work, we should find reevaluations and corresponding eye movement irrespective of the assigned role.

In our second study, we do find the expected coherence shifts. We do also find that participants fixate more often, and longer, on items that support their disposition of the case, whether or not it is induced. But fixations and coherence shifts are essentially uncorrelated. Yet coherence shifts on the one hand, and the number or the duration of fixations on the normative items on the other hand, are almost equally predictive for the decision a participant makes or requests. We thus replicate the unexpected result from Study 1. The mental process of finding the decision (documented with eye tracking) and the process of representing it to its intended audiences (documented with coherence shifts) are disjunct. In conclusion we explore in which ways both processes differ. Coherence shifts result from advocacy. The decision-maker provides a coherent justification for her decision. By contrast in the process of finding the decision, the decision-maker is also attentive to potential counterarguments against the decision she is about to make.

The remainder of the paper is organized as follows: Section II covers the first study, Section III the second. Section IV concludes with discussion.

## **II. Study 1: Eye Gaze as a Reflection of the Ultimate Strength of an Argument?**

### **1. Design**

In our experiment, we give participants a stylized torts case. The case is deliberately ambiguous. The case comprises four elements in support of plaintiff’s claim: defendant has caused damage to plaintiff’s property; defendant has violated a rule that one might argue also protects plaintiff’s property: she has been speeding; damage is pecuniary, so that plaintiff can be compensated; if the court decides for plaintiff, others will be deterred from committing comparable torts. Yet the case also comprises four elements that might favor defense. Plaintiff has herself put her property at risk, by constructing a vulnerable oil tank above ground; defendant has acted with the good intention to save the life of a cat; the obligation to pay damages might hit defendant severely; it might have a chilling effect on third parties. These last two elements are manipulated. In the *no conflict* condition, the

amount of damage is moderate, so that these concerns are rather far-fetched. By contrast in the *conflict* condition, defendant would have to file for bankruptcy when obliged to fully compensate plaintiff. Participants are assigned the role of *judge*, counsel for *plaintiff* or counsel for *defendant* and read the following vignette:

A earns 3,000 € / month. He does not own any property that could be seized. In his car, he passes through a village. The speed limit is 30 km/hour. A drives at 55 km/hour. A cat jumps onto the street. A breaks to save the cat. He loses control of the car. The car collides with an aboveground **water** **oil** tank owned by B. Since the tank had been put in place, construction law has changed. Today it would no longer be permissible to build the tank aboveground. An expert testifies that A would not have lost control of the car had he obeyed the speed limit. A's insurance does not cover damage inflicted on aboveground objects off the street.

Repairing the tank costs 1,500 €. **There is a risk of groundwater contamination. This is why the police obliges B to dredge the ground, at the cost of 43,500 €.**<sup>3</sup> B sues A, and asks for **1,500 €** **45,000 €**.

Plaintiff demands full compensation. Defendant argues that at most partial compensation (**500 €** **15,000 €**) would be justified.

Treatment variations are colour-coded; in the experiment, the manipulated items have of course not been highlighted. In the *judge* conditions, participants are asked to decide the case. They know that they will also be asked to provide a written justification. In the *plaintiff* and *defendant* conditions, they are asked to indicate how the court should decide, and to plead in favour of their respective client.

In preparation of their decision, participants see the computer screen reproduced in Figure 1. This decision screen is in the spirit of a decision aid. It makes an offer for structuring the normative thinking about the case at hand. There is one column with features of the case supporting plaintiff, and another column with features supporting defendant. The position of the two columns is counterbalanced between subjects. The four rows cover facts regarding the responsibility of plaintiff and defendant, the remedy, and potential effects on third parties. Rows are counterbalanced between subjects as well. Participants know that, while they look at this screen, their eye movements are recorded.

---

3 Under German law, the owner of land can be held responsible if an event on her land puts groundwater at risk. The police are free to ask the owner to take care of the risk, even if the event has been caused by a third party. In that case, it is left to contract or torts liability to indemnify the owner for the cost of obliging with the police order.



|           | Plaintiff    | Defendant     |
|-----------|--------------|---------------|
| Plaintiff | did nothing  | above ground  |
| Defendant | too fast     | saved cat     |
| Verdict   | made whole   | goes bankrupt |
| Others    | more careful | deterred      |

**Figure 1**  
**Decision Screen**

middle column: items in support of plaintiff's claim  
(below referred to as causation, negligence, compensation, deterrence)  
righthand column: items in support of defense  
(below referred to as contributory [negligence], excuse, excessive [deterrence], chilling)  
for the German original, see Appendix

On the next screen, participants are asked to spell out the reasons for their decision or claim. They can write up to one page. On the following screen they are asked for the amount of money they want defendant to pay. Thereafter participants are reminded of the eight features of the case that had been presented on the screen used for tracking their gaze patterns. For each of these concerns, they are asked to indicate, on a Likert scale from 1 (not at all) to 7 (extremely so), how much weight the argument should carry; how difficult they found it to understand the argument; to which degree the argument was ambiguous; to which degree the argument was in conflict with other arguments. Each set of questions is presented to them on a separate screen.<sup>4</sup>

The experiment was run in the computer lab of the Bonn Max Planck Institute. 139 students participated in the experiment, 36 in the *judge no conflict* condition, 34 in the *judge conflict* condition, 35 in the role of *plaintiff*, and 34 in the role of *defendant*. One of the eye trackers did not properly isolate fixations. For the analysis, we only use data from the 116 participants where eye gaze has been properly recorded (30 in the *judge no conflict* condition, 29 in the *judge conflict* condition, 29 in the *plaintiff* condition, at 28 in the *defendant* condition). 71 of these participants indicated their gender as female, 43 as male, 1 as diverse, and 1

4 For exploratory purposes, we had also asked participants to rate the weight they would assign to nine abstract normative concerns that might be related to the case; we measured social value orientation (Murphy & Ackermann, 2014); justice sensitivity via the Justice Sensitivity Inventory (Schmitt, Baumert, Gollwitzer, & Maes, 2010)); personality via the 24-item Brief HEXACO Inventory (De Vries, 2013), intuitive vs. rational decision making style via the Rational and Intuitive Decision Styles Scale (Hamilton, Shih, & Mohammed, 2016) and third-party inequality-inefficiency trade-off (Rahal, Hoefl, & Fiedler, 2020). As we do not need these supplementary measures for the message of the paper, we do not report them. The supplementary data is available at [https://osf.io/nu4qj/?view\\_only=8948b4037f8941dd9f00f0c236bbf1c5](https://osf.io/nu4qj/?view_only=8948b4037f8941dd9f00f0c236bbf1c5).

did not respond to this question. Participants earned an average of 15€ per hour for their participation.

Eye gaze was recorded with binocular remote Eye Tribe trackers at a 60Hz sampling rate following a 9-point calibration procedure. Participants were seated at about 60cm distance from 14" Dell Latitude E5440 laptops with a native resolution of 1366 x 768 pixels, and eye trackers were mounted directly below the screen. We defined two types of areas of interests (AOI) on the grid screens, choosing to maximize AOI size instead of using the small preregistered AOI sizes. Non-target AOIs containing labels were defined as 190 x 170 and 280 x 170 pixels in size. Target AOIs containing arguments associated with each category are defined as 280 x 340 pixels in size. We defined fixations with a 30 pixel tolerance in the summed deviation of points' maximum and minimum coordinates on the x- and y-axes and a minimum duration of 50ms (Salvucci & Goldberg, 2000).

## 2. Hypotheses

The main innovation of our paper is the use of eye tracking as a window into the process of deciding a legal case. This is why our hypotheses focus on the effect of the experimental manipulations on eye gaze.<sup>5</sup> This focus is motivated by the eye-mind hypothesis.<sup>6</sup> Specifically, we expect that cases with normative conflict require participants to access more information than cases without conflict, and that they need more time to process the information (duration of fixations). Recall that the *conflict* manipulation is confined to the "remedy" and "other" items (last two rows in Figure 1). We predict

### Hypothesis 1 (conflict)

- a) In the conflict cases, the duration of fixations on content items, and the number of fixations on these items are higher than in the no-conflict cases.
- b) In the conflict cases, the duration of fixations and the number of fixations on the "remedy" and "other" items is higher than in the no-conflict cases.
- c) In the conflict cases, the duration of fixations and the number of fixations on the "remedy" and "other" items is higher than on the remaining items.

In the adversarial system, the counsel for plaintiff and the counsel for defendant are not only allowed, they are even supposed to be partisan. In the interest of creating a level playing field, either party to the case may expect her counsel to present her respective cause in

---

5 We have preregistered the hypotheses at [https://osf.io/nu4qj/?view\\_only=8948b4037f8941dd9f00f0c236bbf1c5](https://osf.io/nu4qj/?view_only=8948b4037f8941dd9f00f0c236bbf1c5). We also had hypotheses regarding total response time. Results testing these hypotheses are available in \*\*\*. As this paper covers the original and the replication study, for brevity we refrain from reporting these results.

6 For a discussion of the relevant literature, see the introduction.

the most favourable light. The manipulation thus introduces an institutionalised bias. We therefore predict:<sup>7</sup>

### **Hypothesis 2 (counsel)**

- a) The number and the duration of fixations allocated to the content AOIs in the “client” column is higher than to the content AOIs in the “opponent” column.
- b) The pattern of fixations differs between attorneys for the plaintiff and for the defendant.
- c) Attorneys exhibit a smaller number of fixations, and a lower fixation duration, than judges seeing the same case.
- d) The pattern of fixations is more balanced in judges (seeing the same case) than in attorneys.

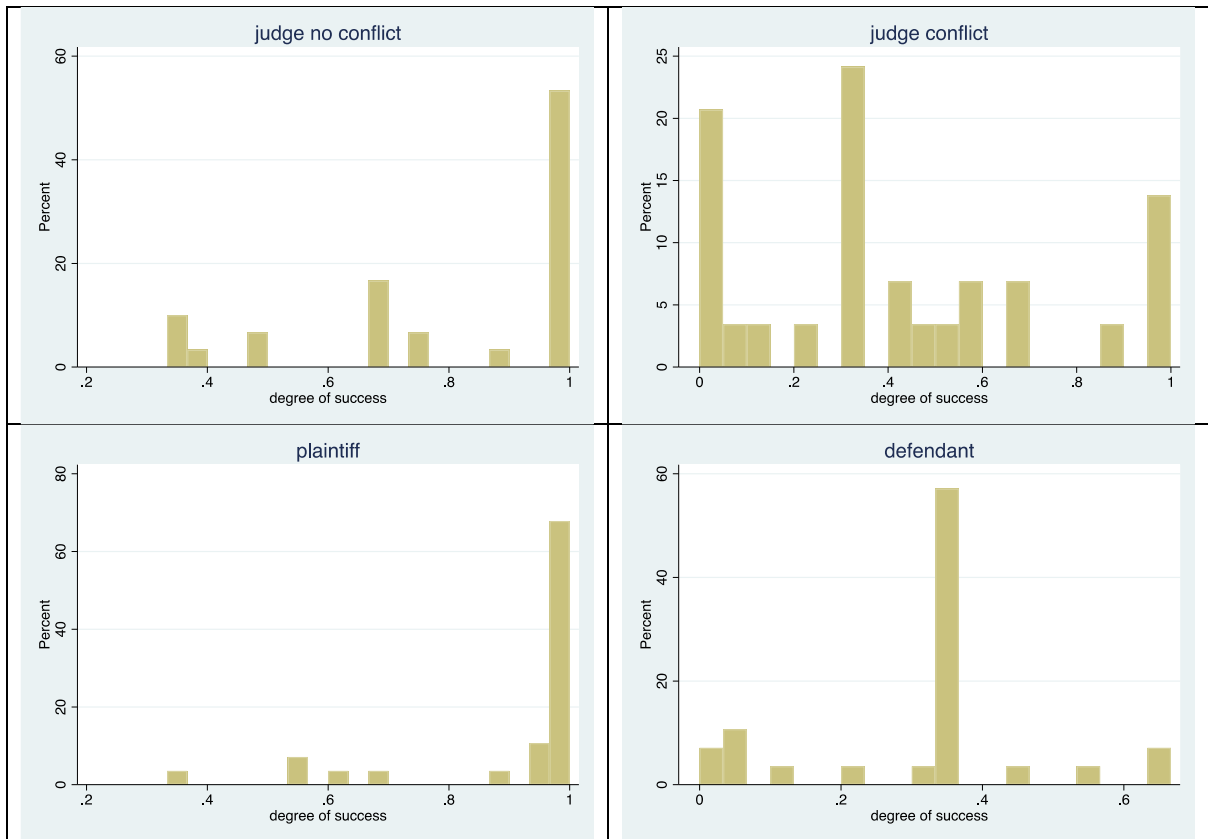
### **3. Results**

Figure 2 shows that our manipulations have worked. Judges assign (much) lower amounts of damage in the *conflict* condition (Mann Whitney,  $N = 59$ ,  $p < .001$ ). Counsels for plaintiff demand substantially higher amounts than judges are willing to grant (Mann Whitney,  $N = 57$ ,  $p < .001$ ). Counsels for defendant do also behave as expected. Descriptively, they plead for even less damages than judges. But this difference is not significant. However to our surprise, none of the hypotheses is supported by the data.<sup>8</sup>

---

7 We also had preregistered the following hypothesis regarding the relationship between eye gaze and self-report: “The higher the importance rating of arguments against the client in the rating task, the smaller the difference in the proportion of attention allocated to the content AOIs in the “client” vs. “opponent” columns”. We discuss this relationship below 4.

8 The detailed results are available from the authors upon request.



**Figure 2**  
**Verdict by Treatment**

One might therefore be concerned that eye tracking is simply not helpful for understanding decision-making in a legal context. In this section we show that this conclusion would be wrong. As the regressions in Table 1 demonstrate, with only information about the number or the duration of fixations, and the condition, we can predict the amount of damages that the participant assigns, requests, or is willing to grant.<sup>9</sup> We can do so with surprisingly high precision. The adjusted  $R^2$  of the regressions in Table 1 is between .819 and .878.<sup>10</sup> Figure 3 visually represents the predictive power of gaze data, by comparing observed choices with the choices predicted when only using gaze data, and treatment.

9 Our explanatory variables are the fraction of all fixations, or the fraction of total duration on one of these 8 content items, on the item in question. Hence by design for each participant the 8 fractions add up to 1. Were we to estimate a regression with a constant, one of the items would drop out. Coefficients would no longer inform us about the relative weight of the item in question for the decision, but about the difference in weight between the omitted item and the remaining items. This is why we estimate models without a constant.

10 The maximum of course being 1.

|                    | number of fixations      |                      | fixation duration        |                      |
|--------------------|--------------------------|----------------------|--------------------------|----------------------|
|                    | model 1                  | model 2              | model 3                  | model 4              |
|                    | conflict<br>manipulation | role<br>manipulation | conflict<br>manipulation | role<br>manipulation |
| causation          | 0.998**<br>(.306)        | .909***<br>(208)     | 1.193**<br>(.362)        | 1.020***<br>(.236)   |
| negligence         | .285<br>(.557)           | .272<br>(.312)       | .063<br>(.640)           | .120<br>(.333)       |
| compensation       | 1.069**<br>(.372)        | .385*<br>(.172)      | .980**<br>(.266)         | .410**<br>(.131)     |
| deterrence         | .836*<br>(.377)          | .464<br>(.278)       | .744*<br>(.309)          | .545*<br>(.243)      |
| contributory       | .827**<br>(.251)         | .486**<br>(.136)     | .941***<br>(.242)        | .465***<br>(.116)    |
| excuse             | .623<br>(.407)           | -.207<br>(.238)      | .583<br>(.442)           | -.366<br>(.279)      |
| excessive          | 1.006**<br>(.304)        | .505**<br>(.183)     | .966***<br>(.202)        | .512***<br>(.136)    |
| chilling           | .615*<br>(.264)          | .421**<br>(.135)     | .541*<br>(.222)          | .414**<br>(.118)     |
| conflict           | -.401***<br>(.083)       |                      | -.301***<br>(.079)       |                      |
| plaintiff          |                          | .484***<br>(.064)    |                          | .474***<br>(.064)    |
| defendant          |                          | -.136*<br>(.064)     |                          | -.137*<br>(.062)     |
| N                  | 59                       | 85                   | 59                       | 85                   |
| RMSE               | .299                     | .230                 | .292                     | .226                 |
| adj.R <sup>2</sup> | .819                     | .873                 | .827                     | .878                 |

**Table 1**  
**Predictive Power of Eye Gaze**

OLS

causation, negligence, compensation, deterrence, contributory, excuse, excessive, chilling:  
shorthands for items on the screen of Figure 1

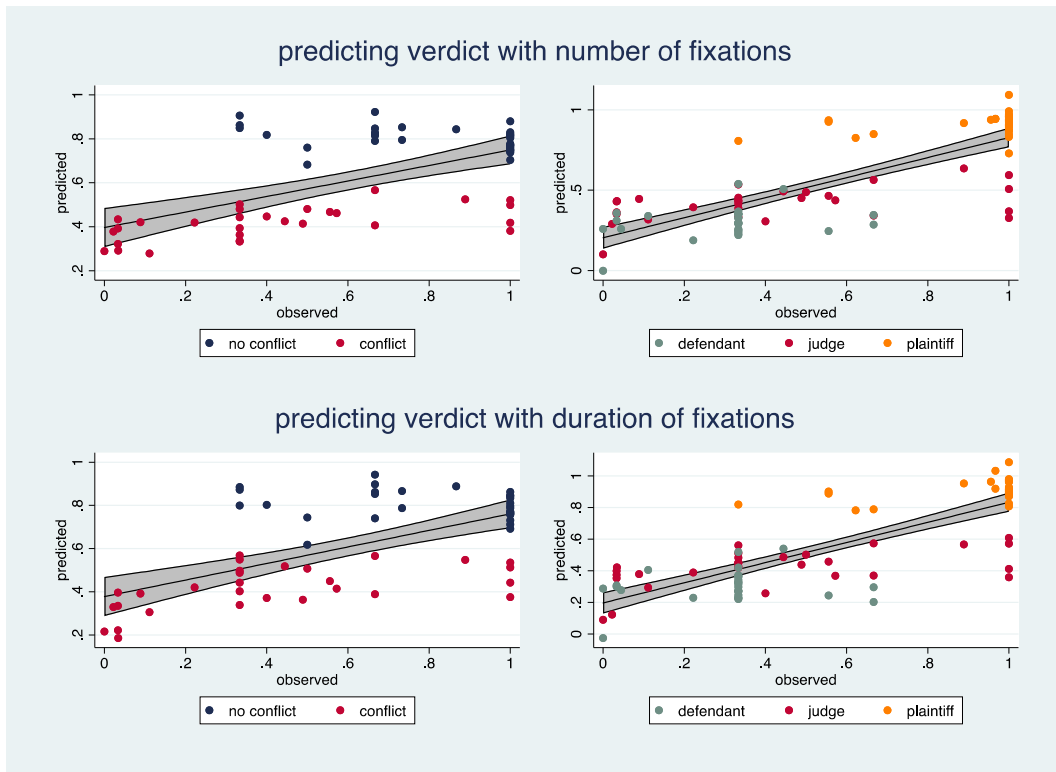
all items represented with proportion, per participant, of fixations on all 8 items / duration of these fixations  
models without constants, as proportions add up to 1 by design

conflict: excessive and chilling items induce normative conflict

plaintiff, defendant: assigned role; reference category in this comparison: judge conflict

one (aggregate) datapoint per participant  
standard errors in parenthesis

\*\*\* p < .001, \*\* p < .01, \* p < .05, + p < .1



**Figure 3**  
**Predictive Power of Eye Gaze**

x-axis: observed damages participant wants to be paid  
 y-axis: predicted damages, exclusively using number or duration of fixations, and treatment

When designing the experiment, we had expected that having participants rate the eight items characterizing the competing normative concerns would help with the interpretation of the number or the duration of fixations. This is why we asked them to rate, on a 7 point Likert scale, the weight they assign to each of the items; the degree of conflict they perceive between the item in question and other items; the difficulty they have in understanding the item; the degree of ambiguity they perceive. Yet the regressions in Table 2 show no evidence that these self-report measures and eye data are related.

|            | model 1             | model 2               |
|------------|---------------------|-----------------------|
|            | number of fixations | duration of fixations |
| weight     | -.0018<br>(.0024)   | -.0021<br>(.0026)     |
| conflict   | .0026<br>(.0024)    | .0026<br>(.0026)      |
| difficulty | -.0015<br>(.0037)   | -.0013<br>(.0041)     |
| ambiguity  | .0029<br>(.0034)    | .0028<br>(.0038)      |
| cons       | .1208***<br>(.0137) | .1221***<br>(.0151)   |
| N obs      | 928                 | 928                   |
| N uid      | 116                 | 116                   |

**Table 2**  
**Interpretation of Eye Gaze**

linear with participant random effect  
dvs: relative number of fixations and relative duration of fixations on one of 8 content items  
ivs: stated weight, degree of conflict, difficulty or ambiguity of item in question  
standard errors in parenthesis  
\*\*\* p < .001

#### 4. Discussion

The main result of Study 1 is unexpected: eye gaze and the explicit evaluation of the probative items are uncorrelated. Eye gaze has however high predictive power. If one knows on which items the participant has looked most often, or longest, one is able to predict her choice with great precision. For the eye-mind hypothesis in a legal context, the evidence is thus mixed. Gaze data is a window into legal decision-making. Otherwise it could not have predictive power. But this mental process is patently not guided by the considerations that motivated our hypotheses. When participants find the solution to the case, they do not predominantly focus on the items that support the decision in the making, or that define the normative conflict.

### III. Study 2: Eye Gaze as a Reflection of Coherence Shifts?

#### 1. Motivation

A non-result does not prove the absence of the expected effect. Ultimately, the non-result is only exploratory. It can help generate a hypothesis. This is what motivates our second study. It has two related purposes. We want to check whether the main result from the first study replicates. And we want to better understand why the explicit ratings of the items and eye gaze were unrelated in study 1.

To the latter end, we have opted for a conceptual, not a literal replication. The added twist of the design is motivated by a psychological theory of intuitive decision-making, the theory of parallel constraint satisfaction (Holyoak & Simon, 1999; Simon, Krawczyk, & Holyoak, 2004). This theory has been specifically developed with an eye to explaining judicial decision-making (Engel et al., 2020; Glöckner & Engel, 2013; Robbennolt, 2005; Simon, 1998, 2004). The theory assumes that automatically spreading activation leads to constructing the most coherent interpretation, taking all constraints into account. In a legal case, the constraints consist of seemingly conflicting pieces of evidence or, in our setting, of seemingly conflicting normative concerns.

The theory posits that inputs are increasingly polarized (Simon, 2004, p. 523). This prediction has been translated into an experimental design. Participants first evaluate a set of items in isolation, and without being tasked to make a decision that takes all items into account. Thereafter they are assigned to the decision-making task. After they have registered their decision, they are once more asked to evaluate the same set of items. If this second set of evaluations systematically differs from the first, this is taken as evidence in support of parallel constraint satisfaction. The effect typically obtains without participants even noticing. The theory calls the effect a “coherence shift” (Holyoak & Simon, 1999). The shift has been observed for various tasks (Brownstein, Read, & Simon, 2004; Glöckner, Betsch, & Schindler, 2010; Russo, Meloy, & Medvec, 1998), including legal decision-making (Carlson & Russo, 2001; Engel & Glöckner, 2013; Simon, 2004).

Our second study combines this experimental paradigm with eye tracking. The traditional experimental paradigm infers parallel constraint satisfaction from the comparison of the ex ante evaluation (before disclosing the task) with the ex post evaluation (after completing the decision-making task). Eye tracking provides evidence “ex interim”, directly during the process of forming the coherent interpretation of the items. Eye tracking has occasionally been combined with the experimental parallel constraint satisfaction paradigm (Fiedler & Glöckner, 2012; Fiedler, Glöckner, Nicklisch, & Dickert, 2013; Glöckner & Herbold, 2011), but – to the best of our knowledge – not for legal decision-making. The promise of pursuing this combination lies in resolving the apparent disconnect between eye tracking and the explicit ex post evaluations that we have observed in study 1. In study 2, we test whether the disconnect results from parallel constraint satisfaction. Eye gaze might reflect the gradual drift of evaluations from the original to the eventual evaluation of the items. If that expectation held true, the number and the duration of fixations on the normative items should be related to the change in explicit evaluations between the first and the second ratings.

## **2. Design**

In Study 2, we use the same case as in Study 1. We do, however, employ the standard design for parallel constraint satisfaction experiments (see e.g. Engel et al., 2020; Glöckner & Engel, 2013; Holyoak & Simon, 1999). Participants initially receive the eight items that may matter for the disposition of the case. But at this point, they are only asked to evaluate each item in isolation, in a task framed as “right or wrong”. Specifically, participants are asked to rate, for each item and on a Likert scale from 1 (not at all) to 7 (extremely so), how much weight



the argument should carry; how difficult they find it to understand the argument; to which degree they perceive the argument to be unclear; to which degree they consider the argument to stand in conflict with other arguments. Argument weight is our central variable of interest.

Most individuals have a preference for consistency (Cialdini, Trost, & Newsom, 1995; Stults & Messe, 1985). Arguably, the perception of being inconsistent is bad for self-esteem (Underwood & Moore, 1981). This motivates the standard design of experiments meant to establish coherence shifts (Glöckner et al., 2010; Lee & Holyoak, 2021), on which we build. Participants do not revisit the same items immediately after they have first evaluated them, but only do so after working on distractor tasks. Distractor tasks do not only create mental distance. They also make it difficult for participants to remember the exact ratings they have given to each of the items. The full set of stimulus materials is available on the Open Science Framework (OSF)<sup>11</sup>.

Right before being confronted with the eight items for the second time, participants are assigned to be either judges or attorneys for plaintiff. They are told that, after they have looked at the decision screen as long as they deem appropriate, they will be asked for their decision (as judges) or for the amount they are claiming (as attorneys), and will be requested to sketch their reasons. We only use the “conflict” version of the case, which is stacked against plaintiff. As this is where the normative conflict is pronounced, we only combine the neutral role of a judge with the biasing role of counsel for plaintiff.

Participants are shown the decision screen of Figure 1. While they reflect upon their decision, their fixations on any one of the eight items are recorded, using the same eye tracking equipment as in Study 1. Participants are aware that eye-trackers are running during the entirety of the study. After making their decision, participants are asked to evaluate the eight normative items for the second time.

In an online behavioural pilot study without eye-tracking<sup>12</sup>, we had tested the materials and found that participants systematically changed their evaluations between the first and second evaluation stages.<sup>13</sup>

The experiment has been preregistered.<sup>14</sup> 119 students of Bonn University with various majors participated in the experiment. Participants received 15€ flat for participating in this study. We excluded data from trials where more than 50% of fixations were outside of the AOIs we had defined in the preregistration (2.98% of the data). In addition, we excluded data from 8 participants whose fixation patterns appeared spurious upon inspection<sup>15</sup>. For three participants, we excluded the data because so few data were collected (total duration < 10 seconds) that a malfunction of the stimulus presentation software seems to have occurred. For another two participants the total number of recorded fixations is below 10, and for one

---

11 [https://osf.io/hmxsw/?view\\_only=63283573b7f54c1fa70b6ed69b6e59ab](https://osf.io/hmxsw/?view_only=63283573b7f54c1fa70b6ed69b6e59ab).

12 Data and materials are available on the OSF.

13 Graphs representing coherence shifts on the weight items is posted on the OSF.

14 [https://osf.io/hmxsw/?view\\_only=63283573b7f54c1fa70b6ed69b6e59ab](https://osf.io/hmxsw/?view_only=63283573b7f54c1fa70b6ed69b6e59ab).

15 Excluding these participants did not substantially change the results reported (see additional analyses without excluding these participants available on OSF).

the total duration is below 10 seconds. For three additional participants, the recorded duration was implausibly high (total duration > 100 seconds), leading us to worry that the participant had trouble completing the task. Of the remaining participants, 51 were assigned to the role of attorney for plaintiff, and 60 to the role of judge.

### 3. Hypotheses

The design of Study 2 is motivated by the theory of parallel constraint satisfaction. The design encompasses the traditional technique for testing the theory. By having participants evaluate the eight items twice, we are able to observe coherence shifts. We expect shifts to obtain if the assigned role biases the outcome (b), but also if participants are unconstrained in their decision, as they assume the role of judge, but grant full compensation (a). The latter expectation follows from the fact that Study 2 uses the “conflict” version of the case, which is stacked against plaintiff.

**H<sub>1</sub> coherence shifts** a) Participants who grant or request full damages increase the declared weight of items in favour of plaintiff, and they decrease the declared weight of items in favour of defendant.

b) Participants who have been assigned to represent plaintiff increase the declared weight of items in favour of plaintiff, and they decrease the declared weight of items in favour of defendant.

The second hypothesis builds on the expectation that eye gaze reflects the drift in explicit evaluations, i.e. the coherence shifts, in the making.<sup>16</sup>

**H<sub>2</sub> eye gaze** a) Participants who grant or request full damages spend more time on items in favour of plaintiff, and exhibit a larger amount of fixations on these items. They spend less time on items in favour of plaintiff, and exhibit a smaller amount of fixations on these items.

b) Participants who have been assigned to represent plaintiff spend more time on items in favour of plaintiff, and exhibit a larger amount of fixations on these items. They spend less time on items in favour of plaintiff, and exhibit a smaller amount of fixations on these items.

The eye-mind hypothesis reported in the Introduction interprets the number and the duration of fixations on the visual representation of decision-relevant information as a window into decision-making, and into information processing in particular. This suggests an additional, behavioral motivation of **H<sub>2</sub>**. Ambiguity gives decision-makers room for manoeuvre. If they adjust the evaluation of the items, while keeping all individual evaluations reasonable, they can get two for the price of one. They get their preferred outcome while maintaining the self-image of a responsible and unbiased decision-maker. If their decision, and possibly

---

16 We also had preregistered the following hypothesis regarding decision time: “participants who have been assigned to represent plaintiff decide faster”. This expectation was motivated by the fact that, for attorneys, the outcome is exogenously imposed, which could make it easier to decide.

also the way how they justify it, become known to third parties, the same strategy also helps them to maintain social esteem.

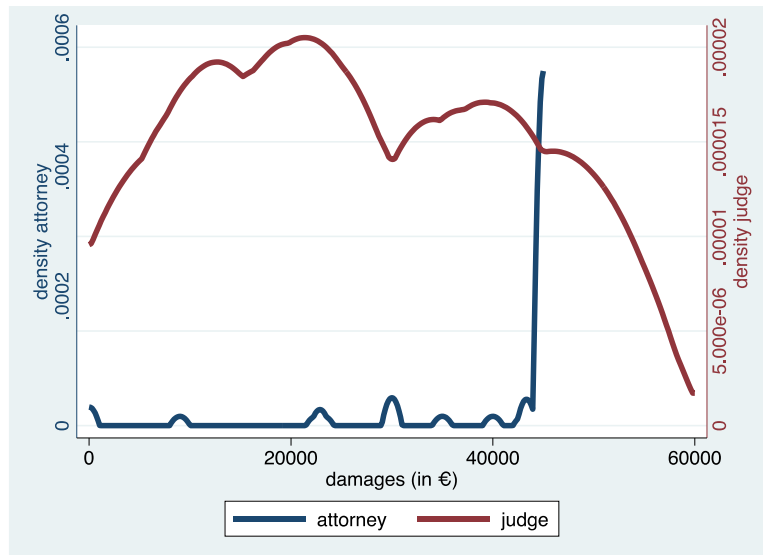
This strategy has been frequently observed as motivated reasoning (Bolsen & Palm, 2019; Epley & Gilovich, 2016; Kunda, 1990). One technique is downgrading the probative value of conflicting evidence (Ditto & Lopez, 1992; Lord, Ross, & Lepper, 1979; Taber & Lodge, 2006). The same technique could also be used for downweighting conflicting normative concerns. Ambiguity provides scope for motivated reasoning (Dieckmann, Gregory, Peters, & Hartman, 2017). Ambiguity specifically allows to serve the goal of being perceived as an accurate, unbiased decision-maker, while at the same time furthering one's preferred substantive interests or ideological position (Bayes, Druckman, Goods, & Molden, 2020; Druckman & McGrath, 2019). Motivated reasoning also matters for legal decision-making (Feldman, 2018). It for instance explains why there is more litigation than would be in the best interest of the parties (Babcock & Loewenstein, 1997). Now the case (in the "conflict" version) is stacked against plaintiff. This is why, for attorneys, the design of the experiment induces the need for motivated reasoning. It has been demonstrated experimentally that "forced compliance" leads to cognitive adjustments of the desirability of the outcome (Engel & Glöckner, 2013; Festinger & Carlsmith, 1959). This is further support for **H<sub>2b</sub>**.

When they see the case for the first time, participants are implicitly in the role of outside observers. They have no reason to engage with the case, and to cut through the normative knot. This is different when they see the case again. Now they are supposed to decide, and their decision is consequential (albeit only hypothetically). As the ambiguity persists, they objectively have to disappoint expectations that are not illegitimate in the first place. Knowing that they inflict harm on one, or both, of the parties may be taxing for their self-image and for their social image. Yet again, ambiguity may come to their rescue. Participants can assuage bad feelings about hurting one or both of the parties by re-evaluating items.

Such reevaluations are predicted by a prominent psychological theory. It posits that decision-makers dislike cognitive dissonance, and exploit opportunities to bring their decision in line with a plausible interpretation of the decision problem (Brehm, 1956; Festinger, 1957; Harmon-Jones & Mills, 2019; McGrath, 2017). Cognitive dissonance may also be caused by a threat to self-consistency (Aronson, 1968), and by feeling personally responsible for the assessment (Cooper & Fazio, 1984). Critically not only attorneys face this need. So do participants in the role of judges. This provides further support for **H<sub>2a</sub>**.

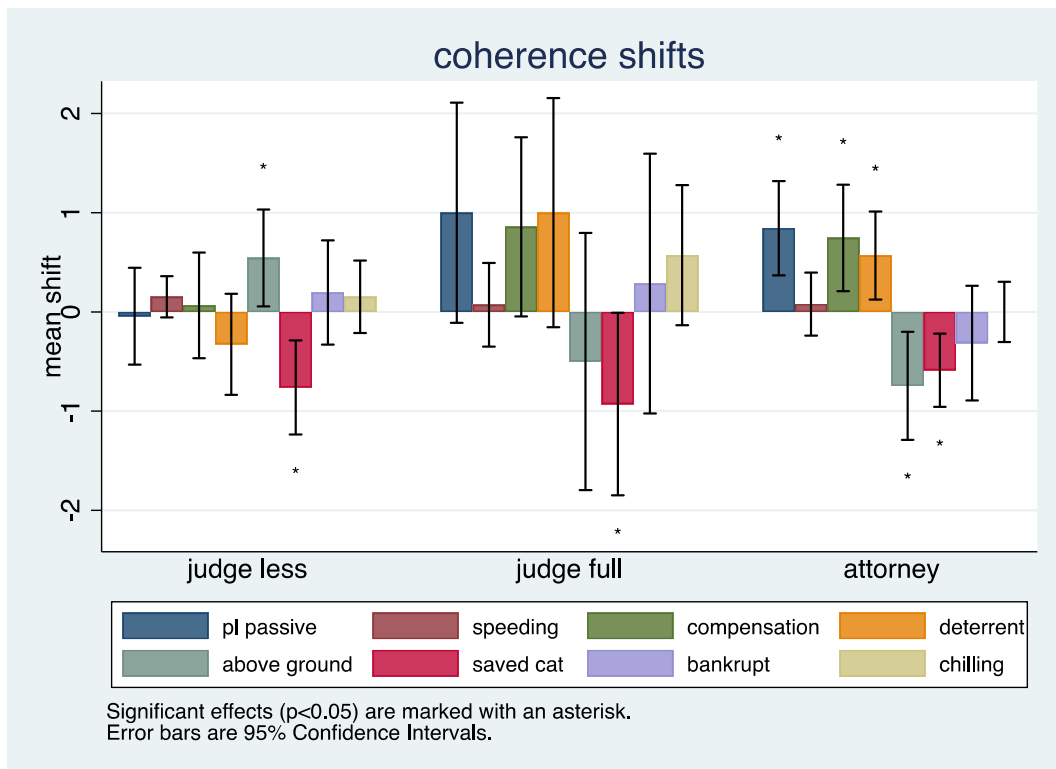
#### **4. Results**

We again start with a manipulation check. As Figure 4 shows, participants assigned to represent plaintiff have predominantly asked for full compensation, while the rulings of participants in the role of judge are spread out, corresponding to the pronounced normative ambiguity induced by the design of the case.



**Figure 4**  
**Damages by Assigned Role**  
 y-axes: densities  
 one outlier (judge, 2.155.022.121€) removed for visibility

To test hypothesis  $H_1$ , in Figure 5 we report coherence shifts, i.e. changes in the evaluation of the eight items, conditional on the decision, and on the assigned role. Participants do indeed change the evaluation of the items between their first and the second exposure to the case. Descriptively, changes are as expected: if they ask for or grant full compensation, several of the items supporting this decision are upweighted, and several of the items speaking against this outcome are downweighted. By contrast, judges who grant less than full compensation adjust their evaluations considerably less.



**Figure 5**  
**Coherence Shifts**

Changes in stated normative weight on any of the 8 items  
on a scale from 1 to 7  
judge less: granted less than full compensation (45.000€)

Table 3 provides statistical tests. From each participant we have two evaluations per item. This complexity is removed by working with the difference in evaluations. Yet moreover from each participant we can calculate this shift for each of 8 items. These shifts are not independent. We capture this dependence by running multivariate (structural) models. We want to learn whether the assignment of a role and the request to decide the case have affected evaluations. Hence, we want to know whether the shift variables are significantly different from zero. In Lines 2-4 we test this, separately for each of the three types of participants, by testing the respective constants against zero.

If judges have granted less than full compensation, we only find two significant coherence shifts. In their ex post evaluation, they deem it more important that the tank has been constructed above ground, and they deem it less important that the defendant intended to save the cat. For the remaining two types of participants, we find systematic patterns largely in line with the decision they have taken. If judges have granted full compensation, ex post they declare it more important that plaintiff has been passive, that compensation would make plaintiff whole, and that others are deterred from violating someone's property as a result of speeding. By contrast ex post they deem it less important that the defendant wanted to save the cat. The only (weakly significant) effect that is not in line with the ruling is the upweighting of a possible chilling effect on third parties. The picture is even clearer for attorneys. After having pleaded for their client, they declare it more important that their client has been passive, that compensation will indemnify her, and that others are deterred.

Ex post they declare it less important that the tank is situated above ground, and that the defendant wanted to save the life of the cat.

|               | pl<br>pas-<br>sive | speed-<br>ing  | compen-<br>sation | deter-<br>rent   | above<br>ground   | saved<br>cat           | bank-<br>rupt   | chilling        | N  |
|---------------|--------------------|----------------|-------------------|------------------|-------------------|------------------------|-----------------|-----------------|----|
| judge<br>less | -.043<br>(.247)    | .152<br>(.135) | .065<br>(.265)    | -.326<br>(.246)  | .543*<br>(.271)   | -<br>.761***<br>(.214) | .196<br>(.286)  | .152<br>(.169)  | 46 |
| judge<br>full | 1.000*<br>(.495)   | .071<br>(.188) | .857*<br>(.401)   | 1.000+<br>(.515) | -.500<br>(.578)   | -.929*<br>(.410)       | .286<br>(.584)  | .571+<br>(.315) | 14 |
| attorney      | .843***<br>(.234)  | .078<br>(.156) | .745**<br>(.264)  | .569**<br>(.219) | -.745**<br>(.269) | -.588**<br>(.182)      | -.314<br>(.285) | .000<br>(.149)  | 51 |

**Table 3**  
**Coherence Shifts**

linear structural models (multivariate regressions) with only a constant for each of the dependent variables  
 testing whether the coherence shift is significantly different from zero  
 data from attorneys and those judges who have granted full (or higher) compensation  
 constants from this structural model are identical with Line 4  
 standard errors in parenthesis  
 \*\*\* p < .001, \*\* p < .01, \* p < .05, + p < .1

We thus have clear support for **H<sub>1a</sub>**. Evaluations pre and post differ significantly. We conclude:

**Result 1:** Participants who request or grant full compensation change their explicit evaluations such that evaluations support their decision.

The regressions in Table 4 test the second hypothesis.<sup>17</sup> If the participant grants or requests (at least) full damages, she significantly more often fixates on items that support plaintiff. The corresponding effect on the duration of fixations is, however, only weakly significant (p = .084). If we control for ruling and role, we find a significant effect for attorneys, in the predicted direction, but no effect for judges who grant full compensation. Attorneys do fixate more frequently, and longer, on the items supporting their cause.

<sup>17</sup> We again estimate structural models (multivariate regressions) as the number and the duration of fixations are not independent from each other.

|                    | model 1                                |  | model 2                                |  |
|--------------------|--|--|--|--|
|                    | number of fixations on plaintiff items | duration of fixations on plaintiff items | number of fixations on plaintiff items | duration of fixations on plaintiff items |
| full damages       | .076*<br>(.036)                        | .061+<br>(.036)                          |  |  |
| judge full damages |  |  | .069<br>(.057)                         | .077<br>(.057)                           |
| attorney           |  |  | .094*<br>(.038)                        | .078*<br>(.038)                          |
| cons               | .516***<br>(.025)                      | .526***<br>(.024)                        | .500***<br>(.028)                      | .509***<br>(.027)                        |

**Table 4**  
**Fixations on Plaintiff Items**

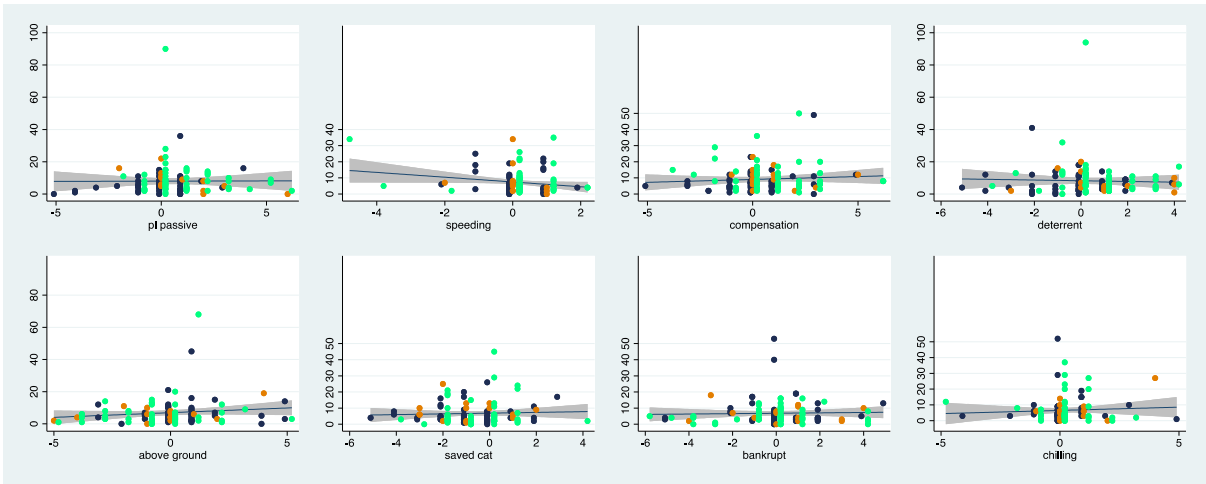
linear structural models (multivariate regressions)  
explaining the fraction of fixations on plaintiff items  
with verdict (model 1) or role and verdict (model 2)  
standard errors in parenthesis  
\*\*\* p < .001, \*\* p < .01, \* p < .05, + p < .1

We thus have qualified support for hypothesis **H<sub>2a</sub>** and **H<sub>2b</sub>** and conclude

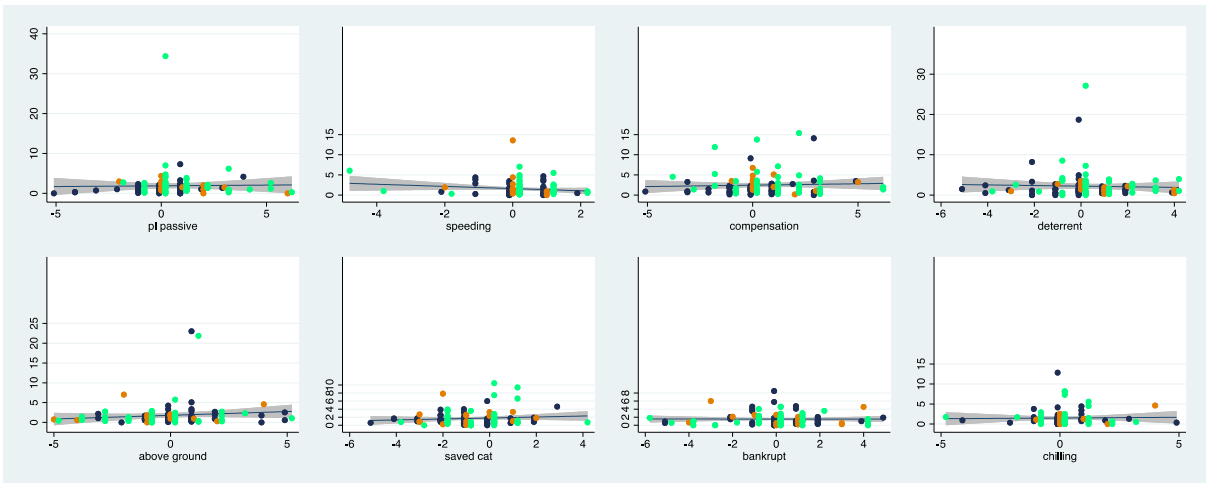
**Result 2:** a) Participants who grant full damages fixate more often on items supporting plaintiff than defendant.

b) Participants who are assigned to the role of attorney for plaintiff fixate more often, and for a longer time, on items supporting plaintiff.

This result shows that participants pay more attention to the items that support their final verdict. In Study 1, we had however found that eye gaze and explicit ex post evaluations are uncorrelated. In Study 2, we have participants evaluate the eight items twice. This makes it possible to study whether eye gaze and the *change* in evaluations are correlated. If true, we would learn that eye tracking portrays the gradual drift of evaluations that ultimately leads to the coherence shift. Yet Figure 6 and Figure 7 show that this is not the case. The figures show that changes in declared weight (i.e. coherence shifts) and eye gaze are essentially uncorrelated, whether we work with the number (Figure 6) or the duration of fixations on the eight normative items (Figure 7). This visual impression is supported by statistical analysis. We only find a significant correlation between the number of fixations on the speeding item and weight on this item, but the correlation is negative.



**Figure 6**  
**(Lack of) Correlation between Change in Declared Weight and Number of Fixations**  
 dots for judge, less than full compensation shifted by -.1  
 dots for attorney shifted by .1, for readability



**Figure 7**  
**(Lack of) Correlation between Change in Declared Weight and Duration of Fixations**  
 dots for judge, less than full compensation shifted by -.1  
 dots for attorney shifted by .1, for readability

We conclude

**Result 3:** The number or the duration of fixations and the shift in explicit evaluations are uncorrelated.

As in Study 1, we want to rule out that gaze data is simply not a good measure of mental process. To that end, we run a horse race of predictive power between coherence shifts on the one hand, and the number or duration of fixations on the other hand. We proceed as follows: our dependent variable is a dummy that is one if the participant has granted (in the role of judge) or requested (in the role of attorney) full (or higher) compensation. We regress this dependent variable on the ex-post evaluations of the eight items, or on the number of fixations on each item, or on the duration of fixations on each item, and control for role. From each of the three regressions we generate predicted values. If the predicted value is



larger than .5, we register the prediction as full compensation, and as less than full compensation otherwise. As Table 5 shows, the predictive power of gaze data is fairly good. We reach 75.44% accuracy, whether we use the number or the duration of fixations. Actually, predictive accuracy when using gaze data is almost as good as the prediction when using explicit evaluations. With the latter input, accuracy is only very slightly higher, 76.58%.<sup>18</sup>

|             | coherence shift |                | number of fixations |                | duration of fixations |                |
|-------------|-----------------|----------------|---------------------|----------------|-----------------------|----------------|
|             | predicted less  | predicted full | predicted less      | predicted full | predicted less        | predicted full |
| actual less | 46              | 13             | 48                  | 14             | 48                    | 14             |
| actual full | 13              | 39             | 11                  | 38             | 11                    | 38             |

**Table 5**  
**Accuracy**

We note

**Result 4:** The number of fixations on each of the normative items, as well as the durations of fixations, predict whether the participant grants or requests full compensation.

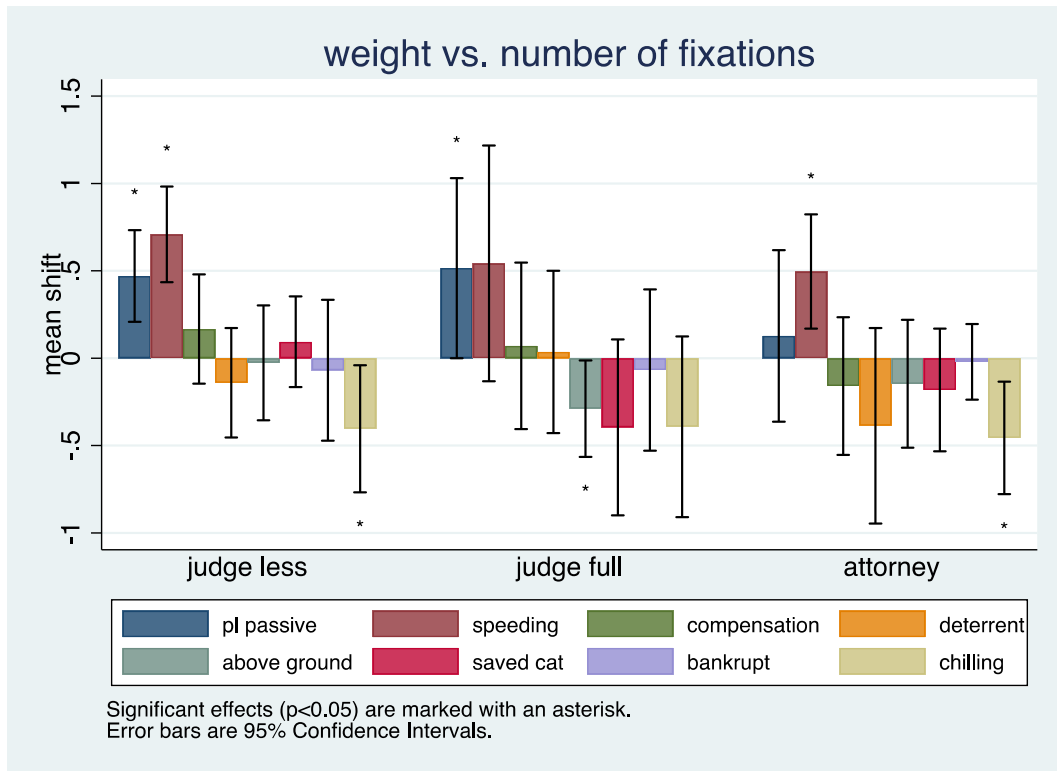
Taking Result 3 and Result 4 together, we further conclude

**Result 5:** The mental process leading to the decision how much to grant or request and the mental representation of this outcome are distinct from each other.

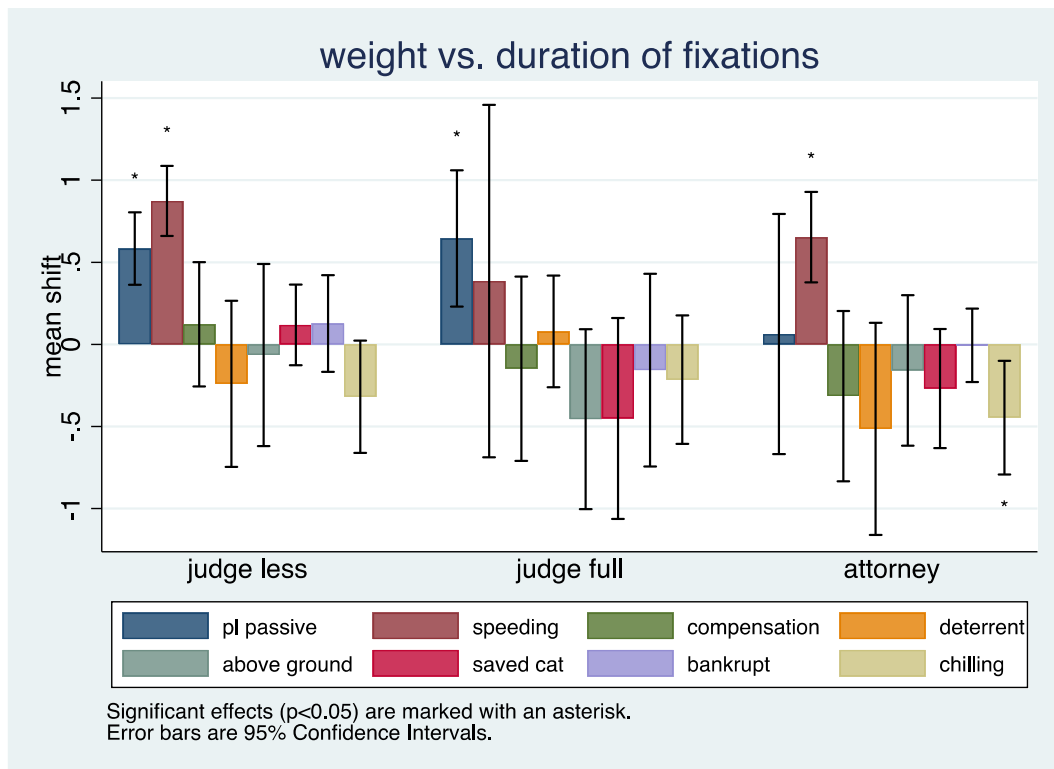
In the final, exploratory step we try to find explanations for the disconnect between eye gaze and coherence shifts. To that end, we compare the explicit weight assigned to an item after having decided with the number or the duration of fixations on the respective item. To that end, we standardize all three measures as fractions of their means, over all participants and items. Figure 8 and Figure 9 show a clear distinction between the context of discovery (eye gaze) and the context of representation (ex post evaluations). The declared weight on the first two items (plaintiff has been passive and therefore not caused the damage on her property; had defendant not speeded, no damage would have been done) is higher than the number of fixations on these items, and their duration. By contrast the declared weight of items supporting defendant's position is smaller than the relative number of fixations on these items, and their duration. These effects are more pronounced for attorneys and for judges who grant full compensation. Hence the explicit downweighting is largely in line with

18 These regressions are available on the OSF. If we withhold role information (i.e. only explain full damages with the set of eight ex post evaluations / number of fixations on either item / duration of fixations on them), unsurprisingly accuracy is lower. We then find 69.30%, 62.28% and 58.77% accuracy, respectively. As we try to separately explain the individual ex post evaluation with either the number or the duration of fixations on the item in question, we do not have to worry that words on the decision screen have different length. In principle, longer words might be harder to understand, which might induce more or longer fixations. But if we normalize the number or the durations of fixations by the comparative length of the word, results, and in particular significance levels, do not change. This is expected, as we only (have to) scale the respective explanatory variable.

the ruling. Participants essentially declare a higher weight for items supporting their chosen outcome, and a lower weight for items in conflict with this outcome. By contrast eye gaze shows that, in the process of finding the solution, participants also have to elaborate on items speaking against the choice they are about to make.



**Figure 8**  
**Coherence Shifts vs. Number of Fixations**  
 Change in stated normative weight - number of fixations on any of the 8 items  
 both scales normalized to difference from the mean over all participants and items  
 judge less: granted less than full compensation (45.000€)



**Figure 9**  
**Coherence Shift vs. Duration of Fixations**  
 Change in stated normative weight – duration of fixations on any of the 8 items  
 both scales normalized to difference from the mean over all participants and items  
 judge less: granted less than full compensation (45.000€)

We conclude

**Result 6:** In the process of finding their decisions, participants do not elaborate systematically more on normative items speaking in favour of their decision.

This finding might be taken as evidence against the theory of parallel constraint satisfaction. After all, coherence shifts are a standard technique to prove the theory, but coherence shifts and the number or duration of fixations on the items are disconnected. Yet we consider this conclusion to be premature. The theory of parallel constraint satisfaction posits that an ill-defined problem is made tractable by exploiting the ambiguity. A consistent representation is found by gradually altering the impact of individual items. Critically this process of reconfiguring the representation does not only require finding enough support for the final outcome. It is equally important to ascertain that potential counterarguments can be moved out of the way. Our finding is consistent with the theory once we allow for the possibility that the process of reconfiguring the evidence and the representation of the outcome found this way are separate mental activities. We have found empirical support for the epistemic distinction of a context of discovery (made visible with the help of eye tracking) and the context of representation (made visible with a change in the explicit evaluation of the evidence, i.e. in the coherence shifts).

#### **IV. Conclusion: Eye Gaze as a Window into the Mental Process of Deciding an Ambiguous Legal Case**

The judiciary routinely handles cases although the facts are contested and the normative assessment is disputed. While the parties, unsurprisingly, are not always happy with the outcome, the general sense is that judicial decision-making does a reasonable job. How can it if the input is, at least partly, inconclusive? In this project we bracket uncertainty about the facts and focus on normative ambiguity. We test student participants on the vignette of a torts case. The case is described such that alternative decisions can be justified, but that full compensation is not obviously the appropriate decision.

Our first study is motivated by the expectation that the explicit ex post evaluation of the decision inputs reflects the mental process, as observed with eye tracking. This expectation is not borne out by the data. But the number and the duration of fixations on the decision inputs have almost the same predictive power as the explicit evaluations, which can be interpreted as explicit justifications. Our second study capitalizes on a psychological theory about deciding ambiguous problems. The theory of parallel constraint satisfaction posits that decision-makers make an ill-defined problem tractable by subconsciously reevaluating the decision inputs until a coherent representation emerges. We use the corresponding experimental paradigm to show that, indeed, such reevaluations take place. After they have made up their minds, participants evaluate the decision inputs differently, and such that evaluations conform with the disposition of the case. But again we find a disconnect with eye tracking data. Both reevaluations (coherence shifts, as this literature calls them) and the number or the duration of fixations on the decision inputs predict decisions, with approximately the same accuracy. But they are essentially uncorrelated. We do not see this as evidence against the theory of parallel constraint satisfaction. Our data rather demonstrates that, indeed, finding the decision and representing it are distinct mental activities, which arguably are both, but independently, results of parallel constraint satisfaction.

Our findings provide an explanation for the seeming puzzle of legal decision-making. Many legal cases are at least ambiguous, if not even technically unsolvable, as there is no theory for trading the competing concerns against each other. Parallel constraint satisfaction explains why judges do not decide the large majority of cases by burden of proof. The mental process reconfigures the inputs, within plausibility constraints, until the problem becomes tractable. Coherence shifts demonstrate that the reasons judges give for their decisions are not a mirror image of this mental process. Rather they must be understood as acts of persuasion.

We have started this paper with the claim: as a matter of practice, there is no way around balancing. Now that we better understand the mental mechanics of balancing, we are in a better position to evaluate potential normative concerns. The mental process is one of constraint satisfaction. Legal policymakers do therefore not have to worry about arbitrary judicial decisions, or openly partisan decisions. Not only will such clear cases of the abuse of judicial powers be easy to identify. In this respect, written reasons are likely to be helpful. If there is no convincing way to justify the outcome, given the applicable legal rules, very likely

the judge has overstepped the institutional boundaries. But the typical judge will avoid such blatant violations of her judicial duties in the first place, if only to preserve self-image and social esteem. Concern should focus on more subtle effects. What happens within the boundaries of plausibility?

The law has not been inattentive to the concern. It goes a long way to secure impartiality. As a rule, judges are required to justify their decisions, usually even in writing (for the normative debate see Engel, 2007). The competent judge must have been determined before the case reaches the court. If an outsider might suspect the judge to be biased, the judge is recused. If one of the parties worries that the judge has not properly assessed the normative weight of competing concerns, she may appeal the case, to list only the most important safeguards. Still, after the fact, other authorities can only reconstruct how the competent judge has balanced the competing features of the case, or one can have a new judge redo the task. Knowing the underlying mental process should help the law on two fronts: the law is alerted to the scope for the subconscious mental reconfiguration of the decision inputs. And with designs analogous to the present series of experiments, potential interventions can be tested for their effectiveness. Eye tracking is likely to provide credible evidence, precisely because eye movements are not under direct conscious control.

The present paper is meant to open the debate over these foundational issues for the law as a discipline, and for the law as a technique for governing society. In so doing, it demonstrates that eye tracking is a very promising, and equally powerful, but as yet untapped resource for legal research.

## References

- Aleinikoff, T. A. (1987). Constitutional law in the Age of Balancing. *Yale Law Journal*, 96(5), 943-1005.
- Alexy, R. (2003). Constitutional Rights, Balancing, and Rationality. *Ratio Juris*, 16(2), 131-140.
- Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing Simple Choices by Manipulating Relative Visual Attention. *Judgment and Decision Making*, 3(5), 396-403.
- Aronson, E. (1968). Dissonance Theory. Progress and Problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosnberg & P. H. Tannenbaum (Eds.), *Theories of Cognitive Consistency. A Sourcebook* (pp. 5-27). Chicago: Rand.
- Babcock, L., & Loewenstein, G. (1997). Explaining Bargaining Impasse. The Role of Self-Serving Biases. *Journal of Economic Perspectives*, 11, 109-126.
- Bayes, R., Druckman, J. N., Goods, A., & Molden, D. C. (2020). When and How Different Motives Can Drive Motivated political Reasoning. *Political Psychology*, 41(5), 1031-1052.
- Becker, S. I. (2011). Determinants of Dwell Time in Visual Search. Similarity or Perceptual Difficulty? *PloS one*, 6(3), e17740.
- Bolsen, T., & Palm, R. (2019). Motivated Reasoning and Political Decision Making *Oxford Research Encyclopedia of Politics*.
- Brehm, J. W. (1956). Postdecision Changes in the Desirability of Alternatives. *Journal of Abnormal and Social Psychology*, 52(3), 384-389.
- Brownstein, A. L., Read, S. J., & Simon, D. (2004). Bias at the Racetrack. Effects of Individual Expertise and Task Importance on Predecision Reevaluation of Alternatives. *Personality and Social Psychology Bulletin*, 30(7), 891-904.
- Carlson, K. A., & Russo, J. E. (2001). Biased Interpretation of Evidence by Mock Jurors. *Journal of Experimental Psychology: Applied*, 7(2), 91-103.
- Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for Consistency. The Development of a Valid Measure and the Discovery of Surprising Behavioral Implications. *Journal of Personality and Social Psychology*, 69(2), 318-328.
- Cooper, J., & Fazio, R. H. (1984). A New Look at Dissonance Theory. *Advances in Experimental Social Psychology*, 17, 229-266.
- Cooter, R., & Ulen, T. (2012). *Law and Economics* (4th ed.). Upper Saddle River: Prentice Hall.

- Corbetta, M., & Shulman, G. L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature reviews neuroscience*, 3(3), 201-215.
- De Vries, R. E. (2013). The 24-item Brief HEXACO Inventory (BHI). *Journal of Research in Personality*, 47(6), 871-880.
- Dieckmann, N. F., Gregory, R., Peters, E., & Hartman, R. (2017). Seeing What you Want to See. How Imprecise Uncertainty Ranges Enhance Motivated Reasoning. *Risk Analysis*, 37(3), 471-486.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated Skepticism. Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions. *Journal of Personality and Social Psychology*, 63(4), 568-584.
- Dörner, D., & Funke, J. (2017). Complex Problem Solving. What It Is and What It Is Not. *Frontiers in psychology*, 8, 1153.
- Druckman, J. N., & McGrath, M. C. (2019). The Evidence for Motivated Reasoning in Climate Change Preference Formation. *Nature Climate Change*, 9(2), 111-119.
- Eisenberg, T., & Engel, C. (2014). Assuring Civil Damages Adequately Deter. A Public Good Experiment. *Journal of Empirical Legal Studies*, 11, 301-349.
- Eisenberg, T., & Engel, C. (2016). Unpacking Negligence Liability. Experimentally Testing the Governance Effect. *Journal of Empirical Legal Studies*, 13, 116-152.
- Engel, C. (2001). Offene Gemeinwohldefinitionen. *Rechtstheorie*, 32, 23-52.
- Engel, C. (2007). The Psychological Case for Obliging Judges to Write Reasons. In C. Engel & F. Strack (Eds.), *The Impact of Court Procedure on the Psychology of Judicial Decision Making* (pp. 71-109). Baden-Baden: Nomos.
- Engel, C., & Glöckner, A. (2013). Role Induced Bias in Court. An Experiment. *Journal of Behavioral Decision Making*, 26, 272-284.
- Engel, C., Timme, S., & Glöckner, A. (2020). Coherence-Based Reasoning and Order Effects in Legal Judgments. *Psychology, Public Policy, and Law*, \*\*\*, \*\*\*.
- Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, 30(3), 133-140.
- Feldman, Y. (2018). *The Law of Good People. Challenging States' Ability to Regulate Human Behavior*. Cambridge: Cambridge University Press.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row Peterson.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive Consequences of Forced Compliance. *Journal of Abnormal and Social Psychology*, 58, 203-210.

- Fiedler, S., & Glöckner, A. (2012). **The dynamics of decision making in risky choice: An eye-tracking analysis.** *Frontiers in psychology*, 3, 335.
- Fiedler, S., Glöckner, A., & Nicklisch, A. (2012). The Influence of Social Value Orientation on Information Processing in repeated Voluntary Contribution Mechanism (VCM) Games: An Eye-tracking Analysis. In A. Sirigu & A. Innocenti (Eds.), *Neuroscience* (pp. 21-53).
- Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social Value Orientation and Information Search in Social Dilemmas. An Eye-Tracking Analysis. *Organizational Behavior and Human Decision Processes*, 120(2), 272-284.
- Geistfeld, M. (2014). Compensation as a Tort Norm. In J. Oberdiek (Ed.), *Philosophical Foundations of the Law of Torts* (pp. 65-85). Oxford: Oxford University Press.
- Geistfeld, M. A. (2011). The Coherence of Compensation-Deterrence Theory in Tort Law. *DePaul Law Review*, 61, 383-418.
- Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence Shifts in Probabilistic Inference Tasks. *Journal of Behavioral Decision Making*, 23, 439-462.
- Glöckner, A., & Engel, C. (2013). Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence Based Reasoning. *Journal of Empirical Legal Studies*, 10, 230-252.
- Glöckner, A., & Herbold, A. K. (2011). An Eye-Tracking Study on Information Processing in Risky Decisions. Evidence for Compensatory Strategies Based on Automatic Processes. *Journal of Behavioral Decision Making*, 24(1), 71-98.
- Goldberg, J. C., Sebok, A. J., & Zipursky, B. C. (2016). *Tort Law: Responsibilities and Redress*: Wolters Kluwer law & business.
- Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental neuropsychology*, 35(1), 1-19.
- Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of personality assessment*, 98(5), 523-535.
- Harmon-Jones, E., & Mills, J. (2019). An Introduction to Cognitive Dissonance Theory and an Overview of Current Perspectives on the Theory. In E. Harmon-Jones (Ed.), *Cognitive Dissonance. Reexamining a Pivotal Theory in Psychology* (pp. 3-24). Washington: American Psychological Association.
- Holyoak, K. J., & Simon, D. (1999). Bidirectional Reasoning in Decision Making by Constraint Satisfaction. *Journal of Experimental Psychology: General*, 128, 1-29.



- Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, Language, and Memory. Bridging Research from the Visual World and Visual Search Paradigms. *Acta Psychologica*, 137(2), 138-150.
- Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading. From Eye Fixations to Comprehension. *Psychological Review*, 87(4), 329-354.
- Kitchner, K. S. (1983). Cognition, Metacognition, and Epistemic Cognition. *Human Development*, 26(4), 222-232.
- Krajbich, I., & Dean, M. (2015). How Can Neuroscience Inform Economics? *Current Opinion in Behavioral Sciences*, 5, 51-57.
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108, 480-498.
- Landes, W. M., & Posner, R. A. (1987). *The Economic Structure of Tort Law*: Harvard University Press.
- Lee, D. G., & Holyoak, K. J. (2021). Coherence Shifts in Attribute Evaluations. *Decision*, 8, 257-276.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization. The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109.
- Luhmann, N. (1966). *Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung*. Berlin: Duncker & Humblot.
- Lynch, C. F., Ashley, K. D., Aleven, V., & Pinkwart, N. (2006). *Defining Ill-Defined Domains. A Literature Survey*. Paper presented at the Intelligent Tutoring Systems 2006.
- McGrath, A. (2017). Dealing with Dissonance. A Review of Cognitive Dissonance Reduction. *Social and Personality Psychology Compass*, 11(12), e12362.
- Meißner, M., & Oll, J. (2019). The Promise of Eye-Tracking Methodology in Organizational Research. A Taxonomy, Review, and Future Avenues. *Organizational Research Methods*, 22(2), 590-617.
- Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: Theoretical and Measurement Issues in the Study of Social Preferences. *Personality and Social Psychology Review*, 18, 13-41.
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know. Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231-259.
- Oberdiek, J. (2008). Philosophical Issues in Tort Law. *Philosophy Compass*, 3(4), 734-748.

- Oberdiek, J. (2014). *Philosophical Foundations of the Law of Torts*: Oxford University Press.
- Orquin, J. L., & Loose, S. M. (2013). Attention and Choice. A Review on Eye Movements in Decision Making. *Acta Psychologica*, 144(1), 190-206.
- Petersen, N. (2017). *Proportionality and Judicial Activism: Fundamental Rights Adjudication in Canada, Germany and South Africa*: Cambridge University Press.
- Pildes, R. H. (1993). Avoiding Balancing. The Role of Exclusionary Reasons in Constitutional Law. *Hastings Law Journal*, 45, 711-752.
- Popper, K. R. (1935). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien,: J. Springer.
- Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.
- Posner, R. A. (2014). *Economic Analysis of Law* (6th ed.). New York: Aspen Publishers.
- Rahal, R.-M., Hoefft, L., & Fiedler, S. (2020). *Eyes on Morals. An Eye-Tracking Analysis of the Cognitive Processes of Moral Decision Making*.
- Rayner, K. (1998). Eye Movements in Reading and Information Processing. 20 Years of Research. *Psychological Bulletin*, 124(3), 372-422.
- Reitman, W. R. (1965). *Cognition and Thought. An Information Processing Approach*.
- Robbennolt, J. K. (2005). Evaluating Juries by Comparison to Judges. A Benchmark for Judging? *Florida State University Law Review*, 32, 469-509.
- Russo, J. E., & Leclerc, F. (1994). An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. *Journal of Consumer Research*, 21(2), 274-290.
- Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional Distortion of Product Information. *Journal Of Marketing Research*, 35(4), 438-452.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying Fixations and Saccades in Eye-tracking Protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71-78.
- Schauer, F. (1995). Giving Reasons. *Stanford Law Review*, 47, 633-659.
- Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory. Factorial Validity, Location in the Personality Facet Space, Demographic Pattern, and Normative Data. *Social Justice Research*, 23(2-3), 211-238.
- Schunn, C. D., McGregor, M. U., & Saner, L. D. (2005). Expertise in Ill-Defined Problem-Solving Domains as Effective Strategy Use. *Memory & cognition*, 33(8), 1377-1387.

- Schwartz, G. (1994). Reality in the Economics of Tort Law. Does Tort Law Really Deter? *UCLA Law Review*, 42, 377-444.
- Schwartz, G. T. (1996). Mixed Theories of Tort Law. Affirming Both Deterrence and Corrective Justice. *Texas Law Review*, 75, 1801-1834.
- Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic Selectivity During Visual Search. The Influence of Central Processing Difficulty. In J. Hyönö, R. Racach & H. Deubel (Eds.), *The Mind's Eye. Cognitive and Applied Aspects of Eye Movement Research* (pp. 65-88): Elsevier.
- Simon, D. (1998). A Psychological Model of Judicial Decision Making. *Rutgers Law Journal*, 30, 1-142.
- Simon, D. (2004). A Third View of the Black Box. Cognitive Coherence in Legal Decision Making. *University of Chicago Law Review*, 71, 511-586.
- Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of Preferences by Constraint Satisfaction. *Psychological Science*, 15, 331-336.
- Stults, D. M., & Messe, L. A. (1985). Behavioral Consistency: The Impact of Public Versus Private Statements of Intentions. *Journal of Social Psychology*, 125, 277-278.
- Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755-769.
- Theeuwes, J. (2010). Top-Down and Bottom-up Control of Visual Selection. *Acta Psychologica*, 135(2), 77-99.
- Underwood, B., & Moore, B. S. (1981). Sources of Behavioral Consistency. *Journal of Personality and Social Psychology*, 40, 780-785.
- Wedel, M., Pieters, R., & van der Lans, R. (2019). Eye Tracking Methodology for Research in Consumer Psychology. In F. R. Kardes, P. M. Herr & N. Schwarz (Eds.), *Handbook of Research Methods in Consumer Psychology* (pp. 276-292): Routledge.
- Welter, M. M., Jaarsveld, S., & Lachmann, T. (2017). Problem Space Matters. The Development of Creativity and Intelligence in Primary School Children. *Creativity Research Journal*, 29(2), 125-132.
- Yantis, S. (2000). Goal-Directed and Stimulus-Driven Determinants of Attentional Control. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes. Attention and performance XVIII* (Vol. 18, pp. 73-103). Cambridge: MIT Press.
- Ziman, J. M. (2000). *Real Science. What it Is, and What it Means*. Cambridge ; New York: Cambridge University Press.

## Appendix

|           | Kläger            | Beklagter     |
|-----------|-------------------|---------------|
| Kläger    | tat nichts        | oberirdisch   |
| Beklagter | zu schnell        | rettete Katze |
| Urteil    | wiederhergestellt | geht bankrott |
| Andere    | vorsichtiger      | abgeschreckt  |

**Figure 10**  
**German Original of Decision Screen**