

**What the Judge Argues is Not What the Judge Thinks  
Eye Tracking Evidence about the Disconnect Between Judicial Decision-Making and  
Judicial Reasoning**

**Abstract**

Lawyers take it for granted that court rulings can normally not be logically derived from first principles. The case requires a willful decision by the person invested with judicial authority. The court must strike a balance between competing, conceptually incompatible normative concerns. We use a combination of behavioral and eye gaze data to investigate the mental mechanism. Without noticing the inconsistency, participants reinterpret normative arguments such that they support their decision. These reinterpretations are not reflected in the frequency or duration of fixations on the competing items presented on a decision screen. However, both explicit reinterpretations and eye gaze predict choices, with about the same accuracy. There are two independent mental effects. Eye gaze is a window into the process that makes the problem tractable, by gradually reinterpreting the arguments. Explicit reinterpretations serve a persuasive purpose. The decision-maker convinces herself, and her intended audience, that her decision is well-founded.

JEL: D01, D81, D91, K13, K40

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

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

## **Eye Tracking Evidence about the Disconnect Between Judicial Decision-Making and Judicial Reasoning**

### **1. Introduction**

Judges are no machines. They do not calculate the outcomes of legal cases, they decide. This voluntary act of a disinterested person is necessary since the typical legal case is ambiguous. In this paper we bracket ambiguity about the facts (on that see Glöckner and Engel 2013, Engel, Timme et al. 2020), and focus on normative ambiguity. Experimental participants are confronted with an ambiguous case. While some features of the case support the claim, others speak against it. Deciding the case requires balancing (cf. Aleinikoff 1987). There is a long-standing normative debate about the desirability of balancing (Pildes 1993, Alexy 2003, Petersen 2017). Yet, for actual judges, there is no way around. But how can they? Generating empirical evidence on the underlying mental process is the topic of the present paper.

There are different options for conceptualizing the mental process. In a rational choice perspective, the decision-maker makes the case tractable by translating the competing normative concerns into arguments of a utility function, and decides for plaintiff if the subjective utility is higher than the perceived utility from rejecting the claim (the key source is Savage 1954). Essentially the decision-maker forces tractability by expressing incompatible normative concerns in an artificial uniform normative currency, utils. By contrast, psychological research suggests that decision-makers exploit ambiguity to their personal advantage. If they have something to gain from the outcome, or at least a predilection for one of the outcomes, they reweight pertinent inputs until their preferred outcome is supported. This mental process is known as motivated reasoning (the key source is Kunda 1990). An alternative psychological theory argues that decision-makers dislike ambiguity in the first place. This research predicts that decision-makers reinterpret inputs such that the case becomes easy to decide. This mental mechanism is called the reduction of cognitive dissonance (the key source is Festinger 1957). A fourth option is conceptualized as parallel constraint satisfaction (a key source is Thagard and Verbeurgt 1998). It argues that reinterpretation is not strategic. It is not a technique employed to achieve an otherwise desired outcome, or to reduce discomfort from a lack of confidence in one's own decision. Rather this theory argues that reinterpreting ambiguous input is the mental mechanism that makes the initially unsolvable problem tractable.

Mental process matters. If subjective expected utility theory gets it right, decision-making is mechanical. If one knows the normative weight assigned to each of the inputs, and the meta-rule for integrating them, one has taken the person of the judge out of the equation. One can oblige the judge to spell out the inputs and their aggregation. Thereafter, everybody, including a court of appeal, can repeat the exercise. Dispute can be narrowed down to the weight assigned to individual items, and to the aggregation rule if needs be. For the most part, judges are indeed obliged to justify their decisions, usually even in writing (for the normative debate see Engel 2007). Experimental research shows that a justification requirement can be

effective in reducing bias (Liu 2018). But other experimental evidence questions the effectiveness of the intervention: role-induced bias persists even when participants are explicitly admonished that a party position might have made it difficult for them to decide neutrally (Spamann 2020).

At any rate, the law is typically not content with a mere justification requirement, but has implemented a number of additional safeguards. 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 interventions. Such safeguards reflect an insight originally developed in analytic philosophy: the context of discovery and the context of representation may fall apart (Popper 1935: 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). Motivated reasoning would not only provide a mental machinery. It would even give the judicial decision-maker two for the price of one: she may take her personally preferred decision, and she may maintain the self-image of impartiality.

If the mental process is better captured by the reduction of cognitive dissonance, the normative concern looms even larger. Even if the decision-maker is perfectly good-natured, and exclusively motivated to fulfil the intentions of the law, her justification is likely to be doctored. One must worry that the written reasons represent the case in a much clearer light than justified by the inputs. This would have clear policy implications. The reasons would only have value as an attempt to persuade the parties and larger audiences that the decision is acceptable. The reasons could not, or at least not to the same degree, be used as a technology for auditing judicial decision-making.

Motivated reasoning and the reduction of cognitive dissonance could conceptually be applied as ex post corrections. The mental process could be such that the decision-maker first finds her decision, and then exploits the degrees of freedom provided by ambiguity to support this decision by an appropriate reevaluation of the inputs. The theory of parallel constraint satisfaction thinks otherwise. It argues that reevaluation is an inseparable feature of the mental process for finding the outcome. It further posits that this is a subconscious process, and that only the outcome is propelled back to conscious cognition (Holyoak and Simon 1999, Simon and Holyoak 2002). If true, this has profound implications for the normative assessment of legal decision-making. The law not only has reason to doubt the debiasing effect of explicit justification. It must live with the insight that legal decision-making can at best partly be made intersubjectively accessible. If the judge herself does not observe the mental process that leads to her decision, outsiders have even less of a chance.

In this paper, we report on an experiment meant to discriminate between these conceptual options. We exploit a paradigm that has frequently been used in the experimental literature on parallel constraint satisfaction, not only, but also in the legal context (Simon 1998, Simon 2004). We have participants first evaluate a set of normative items in a context where they

are mere outside observers. After a series of distractor tasks, they are again presented with the same case, but are now asked to decide, to justify their decision, and to once more evaluate the items.

If we find significant differences in the evaluation (which this literature calls coherence shifts (Simon, Pham et al. 2001)), we have ruled out subjective utility maximization. (About) half of our participants are randomly assigned to the role of judge, while the other half is assigned to the role of attorney for plaintiff. The latter manipulation induces an outcome (full compensation in our torts case). If we find a significant difference in the way how these two groups of participants reevaluate the items, this is evidence in favour of motivated reasoning. If reevaluation does not differ between both groups, this speaks for dissonance reduction.

While participants find their decision, we present the arguments in favour and against full compensation by way of keywords on a screen. We record the number and the duration of fixations on each of the eight items. If we find that gaze data coincides with reevaluations, we have evidence for the mental process that leads to reevaluation. If, by contrast, we find that gaze data and reevaluations are uncorrelated, but gaze data predicts choices, we have evidence for two distinct mental processes: one that makes the ambiguous decision problem tractable by creating a consistent interpretation of the evidence, and one that maintains a positive self image or reduces cognitive dissonance.

The evidence clearly supports this final interpretation: ex ante and ex post evaluations differ substantially. This holds for judges and attorneys. Evaluations change such that they support the decision the participant has made. Ex post evaluations are, however, not explained by the number or the duration of fixations on the respective item. But the predictive power of gaze data is almost the same as the predictive power of ex post evaluations. Gaze data is more balanced over the items. Participants also focus intensely on items speaking against their final decision. Making the ambiguous case tractable by elaborating on the competing items and representing the decision are two distinct mental operations.

The remainder of the paper is organized as follows: in the next section, we introduce the design of the preregistered <sup>1</sup> experiment. In section 3, we derive hypotheses from behavioral theory. In section 4, we report results. Section 5 concludes with discussion.

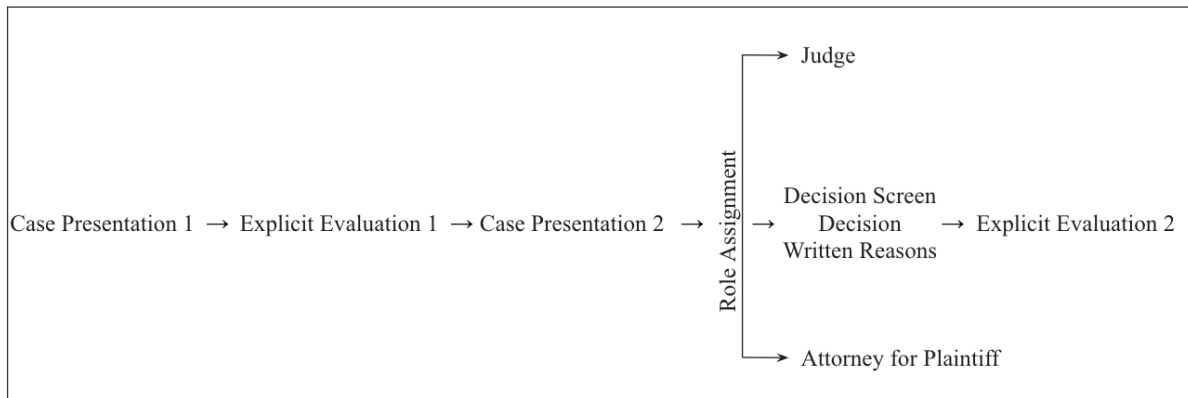
## **2. Design**

Most individuals have a preference for consistency (Stulfs and Messe 1985, Cialdini, Trost et al. 1995). Arguably, the perception of being inconsistent is bad for self-esteem (Underwood and Moore 1981). This motivates the standard design of experiments meant to establish coherence shifts (Glöckner, Betsch et al. 2010, Lee and 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 unlikely that participants remember the exact ratings they have given to each of the items. The full set of stimulus materials is available on the Open Science

---

<sup>1</sup> [https://osf.io/4vbgq/?view\\_only=c94d2ca7494340d8a7654608de041c39](https://osf.io/4vbgq/?view_only=c94d2ca7494340d8a7654608de041c39).

Framework (OSF)<sup>2</sup>. In this section, we focus on the features of the design that are critical for our research question (see Figure 1 for an overview).



*Figure 1*  
*Overview of Procedure*

Participants see and evaluate the same torts case twice. As the experiment is run in Germany, the case is inspired by a classic conflict discussed in the German textbooks for law students (Griesbeck 2019). Under German law, public authorities have a choice whom to address if the public good is in danger. If there is a person who has caused the danger, the authorities may oblige this person to fix the problem. But this person may be unknown or insolvent. If the danger is rooted in a piece of land, the authorities may alternatively also oblige the owner to contain the danger. It is then for the owner to sue the person who has originally caused the damage under the rules of tort law, and to ask for indemnification.

In principle, if the tortfeasor has intentionally or at least negligently caused harm on a passive third-party, under German law she has to pay. But under the rule of comparative negligence (for background see Bohlen 1907, James Jr 1952, De Mot 2013), damages may be reduced if an earlier decision of the victim has made her exceptionally vulnerable. There is also a normative debate over the telos of tort liability (for an indepth discussion see Schwartz 1994, Schwartz 1996, Oberdiek 2008, Oberdiek 2014). Arguably the remedy does not exclusively serve the backward-looking purpose of compensating the victim. It may also be interpreted as a legislative technique for deterring would-be tortfeasors from violating others' property (Landes and Posner 1987). In this forward-looking perspective, it is conceivable to reduce the amount of damages if otherwise the tortfeasor would be deprived of means she needs for living her life. This might also be advisable to avoid a socially undesirable chilling effect, resulting from over-deterrence (De Geest 2012). Some authors have further argued for a subjective element (Ben-Shahar and Porat 2016), which would allow defendant to advance an excuse (Goldberg 2015). Acting with good intentions might be one such excuse. While not spelling out these doctrinal debates, the case exposes participants to the underlying normative conflict.

The case reads as follows:<sup>3</sup>

<sup>2</sup> [https://osf.io/hmxsw/?view\\_only=63283573b7f54c1fa70b6ed69b6e59ab](https://osf.io/hmxsw/?view_only=63283573b7f54c1fa70b6ed69b6e59ab).

<sup>3</sup> For the German original, see the OSF.

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 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 €. B sues A, and asks for 45,000 €.

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

The first time the case is presented, participants read it from a neutral perspective, embedded in a task characterized as containing stories about “right and wrong”. The second time the case is presented, participants are asked to engage with it through the eyes of a judge having to decide the case, or through the eyes of an attorney representing the plaintiff. During the second presentation, we record eye gaze.

To capture the mental process involved in potential reevaluations, we rely on the assumption that attention and eye movements are linked (Just and Carpenter 1980), especially in complex tasks (Rayner 1998), to the extent that "the most active location in working memory will eventually determine the most likely direction of the eye movement at a given point in time" (Huettig, Olivers et al. 2011: 141). We expect that the attention participants pay to a specific normative concern is reflected in the number of fixations to the associated area of interest (AOI) on the decision screen (Orquin and Loose 2013), and in the total duration of fixations on the AOI (Rayner 1998).

To that end, we record the number and duration of fixations on any one of the eight items representing the arguments in the case, displayed on the screen reproduced as Figure 2. Arguments are displayed in two columns, representing the perspectives of the plaintiff and the defendant. In the interest of familiarising participants with the task, they practice using the architecture of the decision screen for another legal case, without having been informed that they will be assigned to the role of plaintiff or judge for the main task. We have added this element to the design to make sure that, during the reevaluation of the target case, we see the decision in the making. We were concerned that, without the additional task, participants would already have made up their minds while we explain the design of the decision screen.

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

*Figure 2*  
*Decision Screen (translated from German, font size increased for legibility)*  
middle column: items in support of plaintiff's claim  
(below referred to as pl passive, speeding, compensation, deterrent)  
righthand column: items in support of defense  
(below referred to as above ground, saved cat, bankrupt, chilling)

Right before being confronted with the target case 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. They are aware that eye-trackers are running during the entirety of the study, i.e., while they work on the distractor tasks and while they are making up their minds in the target task, although eye data was only captured while participants viewed the decisions screen during the second presentation of the target case.

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, in line with the preregistration. 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. As preregistered, 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 and Goldberg 2000). Text positions were counterbalanced between participants.

Regarding explicit evaluations of the arguments, after the initial presentation of the case and again after the second presentation, 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 found it to understand the argument; to which degree they perceived

the argument to be unclear; to which degree the argument stood in conflict with other arguments. Argument weight was the central variable of interest.

In an online behavioural pilot study without eye-tracking<sup>4</sup>, we had tested the materials and found that participants systematically changed their evaluations between the first and reevaluation stages.<sup>5</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>6</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 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

Predictions depend on the assumptions about the underlying mental process. We consider four different ways how decision makers handle the (deliberate) ambiguity of the case.<sup>7</sup>

**Subjective utility.** In the courtroom, the facts are often disputed. Frequently the distribution of possible past events is unknown. Not so rarely, it is not even clear which events might have happened. Then objective probabilities cannot be used, and even the state space is unknown. These are the decision problems for which Savage has developed his theory of decision-making based on subjective estimates of possible events, and of the expected probability of each of these events (Savage 1954), based on concepts originally proposed by Ramsey (1931) and de Finetti (1937). Once the decision-maker has elaborated these inputs, she is back to comparing the expected utility of omission versus commission, or of alternative actions for that matter.

Strictly speaking this is not the problem participants face in our experiment. Facts are undisputed and known with certainty. Their normative evaluation is at stake. But if a decision-maker feels comfortable with the procedure proposed by Savage when facing factual ambiguity, she may use the exact same procedure to tackle normative ambiguity. She may first assess valence (the item in question speaks in favour or against taking the respective

---

<sup>4</sup> Data and materials are available on the OSF.

<sup>5</sup> Graphs representing coherence shifts on the weight items is posted on the OSF.

<sup>6</sup> Excluding these participants did not substantially change the results reported (see additional analyses without excluding these participants available on OSF).

<sup>7</sup> We did not want to preregister four different sets of alternative hypotheses. The preregistered hypotheses are noted in the respective cases, and reported verbatim in the Appendix.

action), and may then assign a normative weight to each of the items. The decision-maker of course needs a meta rule that tells her how to aggregate over the items. The items do not have to be summed up. The decision-maker may allow for substitution (two items pointing into the same direction carry less weight than the sum of their weights, were they observed in isolation) or complementarity (the fact that two items point into the same direction carries even more weight than the sum of the weights assigned to each of the items in isolation). In the spirit of a standard of proof, she may weigh false positive decisions (she fulfils the request although, when being fully informed, she would not have wanted to do so) more heavily than false-negative decisions. Still reasoning is perfectly linear, from the normative weight of each individual item over the decision rule to the outcome.

If this model captures the essence of legal decision-making in the face of normative ambiguity, the initial evaluations of the items are critical. These evaluations should suffice to predict the ruling. Ex ante and ex post evaluations should coincide. The mental processes captured by gaze data should reflect initial evaluations. It should be most difficult for participants to decide in the role of attorney, as now they are supposed to justify a decision that may not be in line with their initial evaluations. Decision difficulty should be reflected in fixation counts, fixation durations, and in decision times (Just and Carpenter 1980, Glöckner and Herbold 2011). This gives us

**H<sub>1</sub> subjective utility:**

- a. Evaluations ex ante and ex post do not differ.
- b. Higher explicit ratings of an item's importance are associated with higher fixation counts and longer fixation durations.
- c. Participants in the role of attorney show higher fixation counts and longer fixation durations.
- d. Participants in the role of attorney take longer to decide.

Savage's theory has spawned a rich empirical literature. It has in particular been shown that probabilities are not weighted linearly (Ellsberg 1961), and that outcomes are evaluated in relation to a reference point (Kahneman and Tversky 1979, Köszegi and Rabin 2006). Both qualifications have been combined in cumulative prospect theory (Tversky and Kahneman 1992). But these qualifications do not affect our hypotheses.

**Motivated reasoning.** This is different with other psychological findings. 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 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 (Kunda 1990, Epley and Gilovich 2016, Bolsen and Palm 2019). One technique is downgrading the probative value of conflicting evidence (Lord, Ross et al. 1979, Ditto and Lopez 1992, Taber and Lodge 2006). The same technique could also be used for downweighting conflicting normative concerns. Ambiguity provides scope for motivated reasoning (Dieckmann, Gregory et al. 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 (Druckman and McGrath 2019, Bayes, Druckman et al. 2020). 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 and Loewenstein 1997).

In the experiment, participants in the role of judges have no reason to tilt the balance. They have no discernible personal interest in the outcome. They have nothing to gain from re-evaluating the items. This is different for participants in the role of attorneys. The design of the experiment forces them to take a decision that may be at variance with their initial assessment of the case. Actually, the case is designed such that there is a normative conflict between legitimate interests of the parties. Very likely full compensation is therefore not the outcome most participants would have deemed appropriate if they had been assigned to be judges. 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 (Festinger and Carlsmith 1959, Engel and Glöckner 2013). If participants in the role of attorney care about preserving self-esteem and social esteem, their mental processes should reflect the reweighting of the normative items that serves this purpose. Finally, as the outcome is induced exogenously, participants in the role of attorney should be faster, as they have no normative conflict to dissolve. This gives us

**H<sub>2</sub> motivated reasoning (preregistered):**

- a. Evaluations ex ante and ex post differ for attorneys, but not for judges.
- b. Higher explicit ratings of an item's importance are associated with higher fixation counts and longer fixation durations.
- c. Participants in the role of attorney fixate more on and spend more time fixating items in favour of plaintiff.
- d. Participants in the role of attorney take less time to decide.

Note that these predictions do not require a specific mental process. The adjustment could be deliberate: the decision-maker constructs a representation of the items that supports her induced assessment. The adjustment could also be subconscious: the decision-maker exploits ambiguity to forge a coherent representation in line with her choice. All we require is the source of the need for adjustment: a conflict between the goal to be perceived as accurate and unbiased, and the competing goal to fulfil the exogenous request.

**Reduction of cognitive dissonance.** 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, McGrath 2017, Harmon-Jones and Mills 2019). Cognitive dissonance may also be caused by a threat to self-consistency (Aronson 1968), and by feeling personally responsible for the assessment (Cooper and Fazio 1984).

Critically not only attorneys face this need. So do participants in the role of judges. As the case is normatively deliberately ambiguous, the need for reevaluation should be more pronounced the higher the damages. Mental process should prepare the desirable re-evaluation. From the vantage point of cognitive dissonance, the decision problem is equally hard for attorneys and judges, so that response times should not differ by assigned role. This gives us

**H<sub>3</sub> reduction of cognitive dissonance (H<sub>3a</sub> - H<sub>3c</sub> preregistered):**

- a. Evaluations ex ante and ex post differ, the more so the higher the damages. 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. Higher explicit ratings of an item's importance are associated with higher fixation counts and longer fixation durations.
- c. 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.
- d. Response time does not differ by assigned role.

**Parallel constraint satisfaction.** The method developed by Savage makes ambiguous decision problems tractable. But the decision-maker has a price to pay. She must be willing to evaluate each item with one and the same normative currency: how intensely does the item in question affect the utility of one of the competing decisions? From a philosophical point of view, this procedure may be questionable. The decision-maker may in particular be hesitant to trade deontological concerns against utilitarian ones (Zamir and Medina 2011). This challenge has been studied prominently, and also empirically, with the famous trolley problems (Thomson 1976, Greene, Cushman et al. 2009, Awad, Dsouza et al. 2018). Even among utilitarian concerns, comparability can appear problematic. The classic illustration is efficiency versus equality (Konow 2005).

Decision problems involving tradeoffs between deontological and utilitarian concerns like the one in our experiment are perfectly common in the judiciary. How is it possible that cases are routinely decided, and that observers for the most part find the decisions acceptable, although the result is not derived from agreed-upon first principles? A psychological theory that capitalises on developments in computer science proposes an answer. Computer scientists have developed algorithms for “constraint satisfaction”, and have enabled them to run in parallel, which is what gives the conceptual approach its name (Yokoo and Hirayama 2000, Rossi, Van Beek et al. 2006, Mostafa, Müller et al. 2015). The theory essentially argues that decision-making is bidirectional (Holyoak and Simon 1999): involving influences not only from input to output, but also from output to input, and possibly in many iterations between input and output, until a solution emerges (Kunda and Thagard 1996, Simon, Pham et al. 2001, Simon, Krawczyk et al. 2004, Glöckner, Betsch et al. 2010, Glöckner, Hilbig et al. 2014). In this

perspective, transforming the decision inputs is what ultimately makes the decision problem tractable (Simon and Holyoak 2002). Respecting all plausibility constraints, the inputs are gradually transformed, until a coherent picture becomes discernible (Thagard and Verbeurgt 1998, Lee and Holyoak 2021). Eye-tracking has been used to make this process visible (Glöckner and Herbold 2011).

This concept of deciding in the face of ambiguity has also been productively used to explain legal decision-making. Earlier work had established that judicial decisions are typically found by way of story telling: the judge or jury member checks whether the stories told by prosecution / plaintiff or defense appear inconsistent with an important piece of evidence (Hastie, Penrod et al. 1983, Pennington and Hastie 1991, Hastie 1993, Pennington and Hastie 1993). Parallel constraint satisfaction provides a fitting model for the underlying mental process (Simon 1998, Simon 2004, Glöckner and Engel 2013, Engel, Timme et al. 2020).

Note the critical difference between parallel constraint satisfaction and motivated reasoning or dissonance reduction. For motivated reasoning and dissonance reduction it suffices to adjust the evaluation of the items after the decision has been taken. It is personal disutility from a disconnect between the materials and the outcome that drives the adjustment. It is of course possible that motivated reasoning and the reduction of cognitive dissonance are mentally executed by way of parallel constraint satisfaction, as claimed by Shultz and Lepper (1996). Yet for motivated reasoning and the reduction of cognitive dissonance, this isn't but one option. By contrast the theory of parallel constraint satisfaction posits that reevaluation is how the decision is found in the first place.

Now the need to bolster self-esteem, and to preserve social esteem, do not disappear. This is why, in the perspective of parallel constraint satisfaction, not only reevaluation is predicted. There is also room for a disconnect between mental process and ex post evaluation. The process of finding the decision (by elaborating on the evidence) and the process of representing the outcome (by transforming the evidence) may fall apart. Finally, in this perspective the assigned role is again immaterial. This gives us

**H<sub>4</sub> parallel constraint satisfaction:**

- a. Evaluations ex ante and ex post differ.
- b. Ex post evaluations on the one hand and the number of fixations on items, or the duration of fixations on these items, on the other hand, may differ.
- c. Fixation counts and fixation durations do not differ by assigned role.
- d. Response time does not differ by assigned role.

#### **4. Results**

We have hypotheses about three dependent variables: a comparison between ex ante and ex post evaluations (a); a comparison between ex post evaluations and eye gaze (b); response times (c). We take them up in turn, and provide data and code on the OSF.

### a) Coherence Shifts

$H_{2a}$  predicts that coherence shifts are more pronounced for attorneys.  $H_{3a}$  predicts that coherence shifts are the more pronounced the higher the damages. We must therefore first check how damage amounts are distributed. For attorneys vs. judges, this also serves as a manipulation check. As Figure 3 shows, the role manipulation has indeed had the intended effect. 38 of 51 attorneys ask for full compensation (45.000€), whereas only 14 of 60 judges grant full compensation (or more).

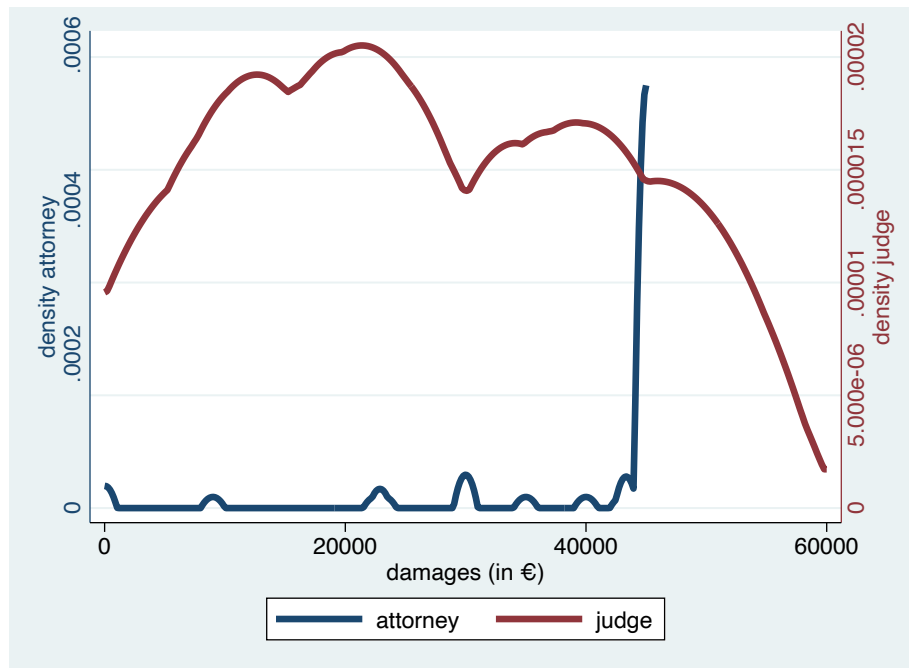
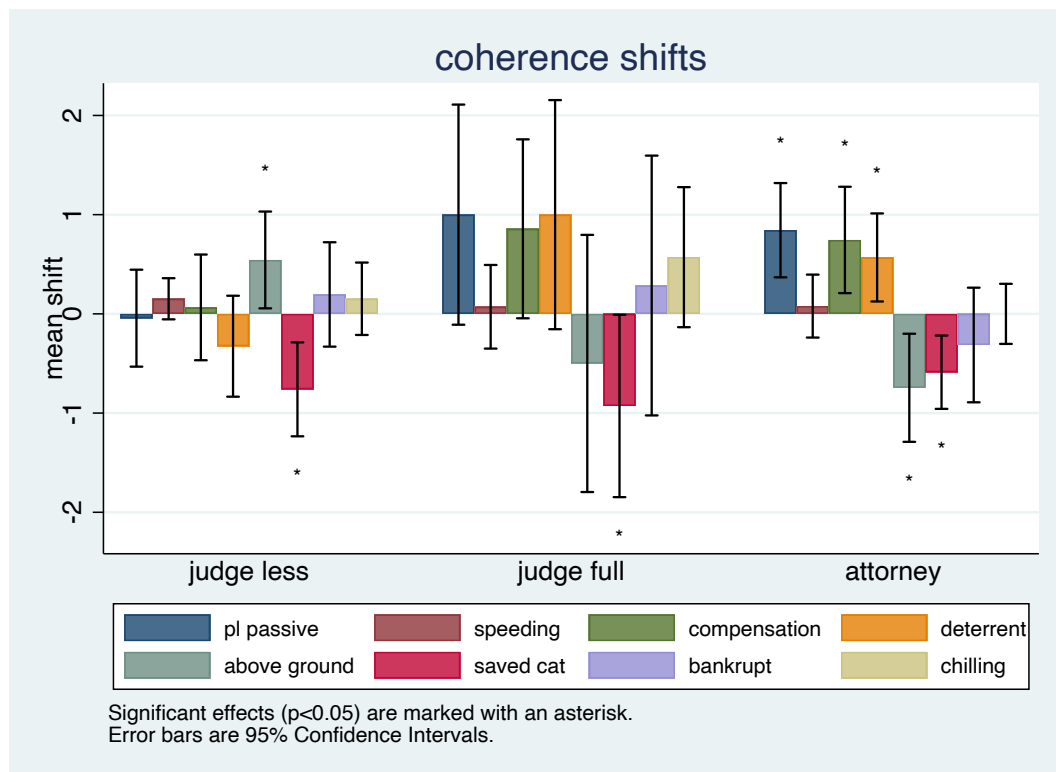


Figure 3  
Damages by Assigned Role  
y-axes: densities  
one outlier (judge, 2.155.022.121€) removed for visibility

As Figure 4 shows, participants do indeed change the evaluation of the items between their first and the second exposure to the case. Descriptively, changes are as expected by  $H_{3a}$ : 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 4*  
*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 1 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.

In Line 5 we present the coefficients from a model that compares attorneys and those judges who have granted full compensation. These coefficients check whether reevaluations by attorneys are significantly different from the reevaluations by judges who have made the same choice. Except for the weakly significant upweighting of the chilling effect, we do not find any significant differences.<sup>8</sup>

	pl passive	speeding	compen- sation	deterrent	above ground	saved cat	bankrupt	chilling	N
judge less	-.043 (.247)	.152 (.135)	.065 (.265)	-.326 (.246)	.543* (.271)	-.761*** (.214)	.196 (.286)	.152 (.169)	46
judge full	1.000* (.495)	.071 (.188)	.857* (.401)	1.000+ (.515)	-.500 (.578)	-.929* (.410)	.286 (.584)	.571+ (.315)	14
attorney	.843*** (.234)	.078 (.156)	.745** (.264)	.569** (.219)	-.745** (.269)	-.588** (.182)	-.314 (.285)	.000 (.149)	51
judge full (vs. attorney)	.157 (.517)	-.007 (.314)	.112 (.547)	.431 (.497)	.245 (.595)	-.340 (.409)	.599 (.624)	.571+ (.329)	65

Table 1  
Coherence Shifts

Lines 2-4: 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

Line 5: coefficients from a linear structural model with the same set of dependent variables, data from attorneys and those judges who have granted full (or higher) compensation constants from this structural model are identical with Line 4

We thus have clear support for **H<sub>3a</sub>**. Evaluations pre and post differ significantly. This speaks against a Savage type of dissolving the ambiguity. Reevaluations by attorneys and by those judges who have made the decision to award full damages do not differ. This speaks against motivated reasoning. We conclude:

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

## b) Parallel constraint satisfaction

If the model by Savage captures the essence of decision making in the face of normative uncertainty, the decision maker must aggregate over the items in favour of either plaintiff or defendant, weighing each item as expressed in the weights when originally seeing these items. If either motivated reasoning or dissonance reduction leads to a reweighting of some of the items, this reevaluation must be mentally prepared. Eye tracking is a window into the mental process that leads to judgement and decision-making. For these reasons **H<sub>1b</sub>**, **H<sub>2b</sub>** and **H<sub>3b</sub>** predict that the number of fixations and their duration on the one hand, and the (for **H<sub>2b</sub>** and **H<sub>3b</sub>** posterior) evaluations are positively correlated. However, Figure 5 and Figure 6

<sup>8</sup> Results look very similar, and significance levels do not change if, for consistency, we constrain attorney data to attorneys who have claimed at least 45000, i.e. full damages (see additional analyses provided on the OSF).

suggest that posterior evaluations (on the respective x-axis) and either the number of fixations or their duration (on the y-axis) are largely uncorrelated: the fitted lines are essentially flat.



Figure 5  
(Lack of) Correlation between Declared Weight and Number of Fixations  
dots for judge, less than full compensation shifted by -.1  
dots for attorney shifted by .1, for readability

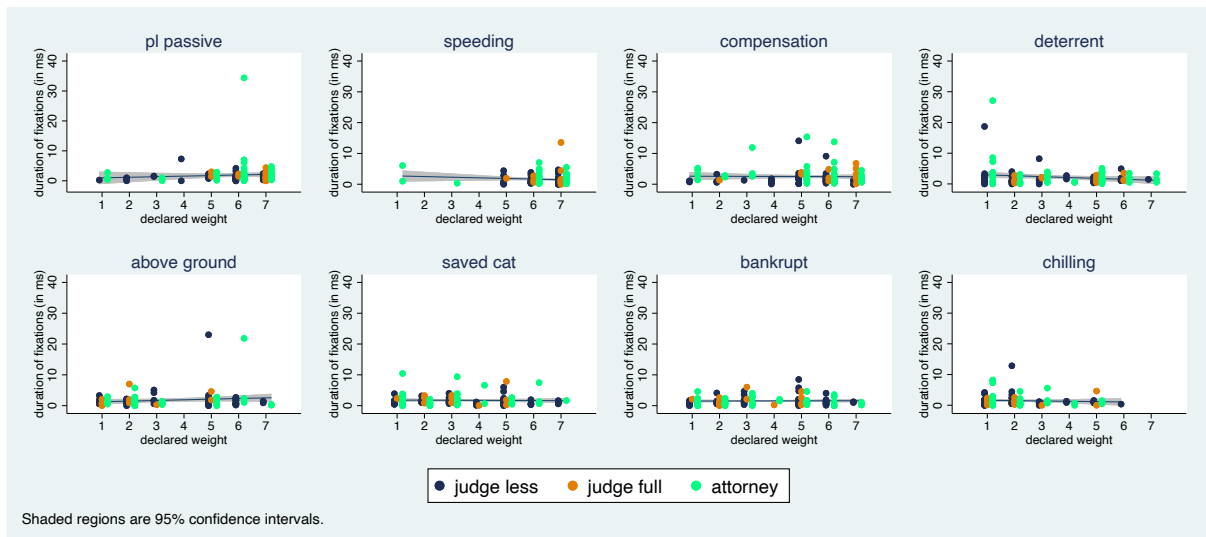


Figure 6  
(Lack of) Correlation between Declared Weight and Duration of Fixations  
dots for judge, less than full compensation shifted by -.1  
dots for attorney shifted by .1, for readability

This visual impression is supported by statistical analysis.<sup>9</sup> If we regress ex post evaluations on the number of fixations on the respective item, at conventional levels we only find a

<sup>9</sup> We again capture dependence at the level of participants by a linear structural model (multivariate regression). All structural models are available on the OSF.

significant effect for the speeding item. Yet the correlation is negative: the more frequently the participant has looked at this item, the more she is likely to evaluate the item as unimportant. We also find the weakly significant effect of the deterrence item, which is also negative. The picture essentially remains the same if we split the statistical analysis by role and ruling. For judges who have granted less than full compensation, the only weakly significant coefficient is again for the speeding item, and it again is negative. For judges who have granted full compensation, the only significant coefficient is for the above ground item (and this coefficient is positive, showing the expected positive correlation). Finally for attorneys, we find significant coefficients for the speeding, compensation and deterrence items, but all coefficients are negative.

If, instead, we try to explain ex-post evaluations with the duration of fixations on the respective item, we find even less significant effects. If we pool the data over role and ruling, we only find a weakly significant, negative coefficient for the deterrence item. If we split the data by role and ruling, we do not find any significant effect for judges, irrespective of their ruling. For attorneys, we find a significant effect at conventional levels for the deterrence item, at a weakly significant effect for the speeding item. But both effects are negative. This gives us

**Result 2:** The number of fixations on normative items, and the duration of fixations, is uncorrelated with the ex post evaluation of the items.

This lack of correlation could result from the fact that gaze data is simply not a good measure of mental process. To rule this explanation out, we run a horse race of predictive power between ex post evaluations 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 2 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, we only find one more true positive, and one more true negative, which yields an accuracy of 77.19%.<sup>10</sup>

---

<sup>10</sup> 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 67.54%, 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.

	weight		number of fixations		duration of fixations	
	predicted less	predicted full	predicted less	predicted full	predicted less	predicted full
actual less	49	13	48	14	48	14
actual full	10	39	11	38	11	38

*Table 2*  
*Accuracy*

We note

**Result 3:** 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 2 and Result 3 together, we further conclude

**Result 4:** 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 the ex post evaluation of the items. To that end, we standardize all three measures as fractions of their means, over all participants and items. Figure 7 and Figure 8 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 happened) 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 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.<sup>11</sup>

---

<sup>11</sup> In Table 3 in the Appendix, we report statistical tests for differences in the standardized declared weight vs. eye gaze, separately for each of the 8 items. In the Appendix, we also present descriptive statistics about the (lack of) correlation between the explicit statement that an item was difficult to ascertain, unclear, or in conflict with other items, compared with either the number or the duration of fixations. In these comparisons too we find practically no significant correlations. The additional regressions showing the lack of correlation are available from the authors upon request.

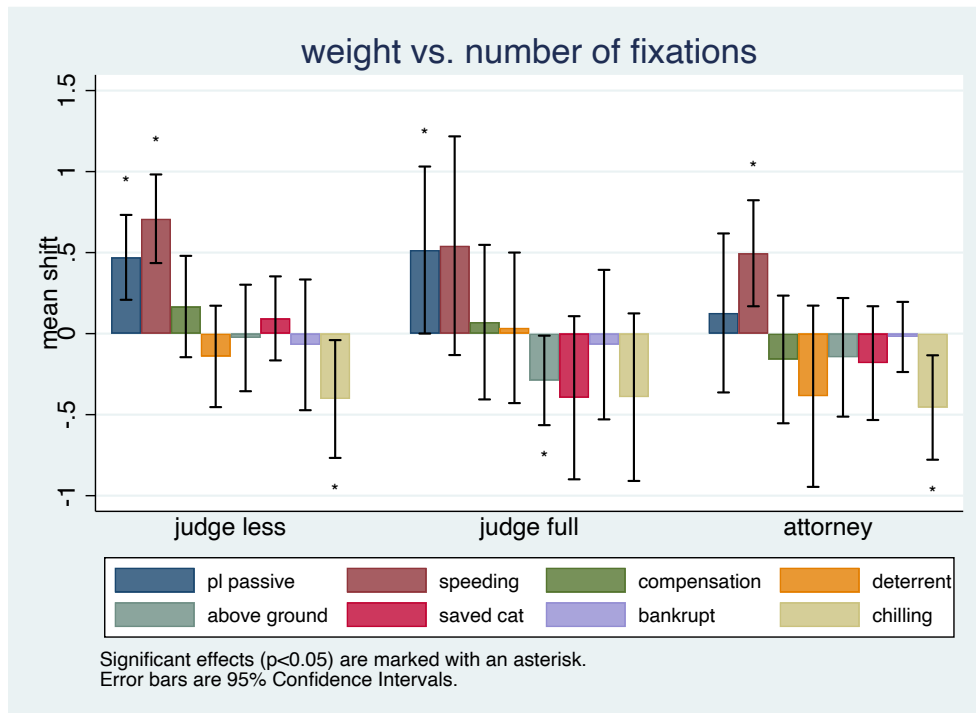


Figure 7

*Ex Post Evaluation vs. Number of Fixations*

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€)

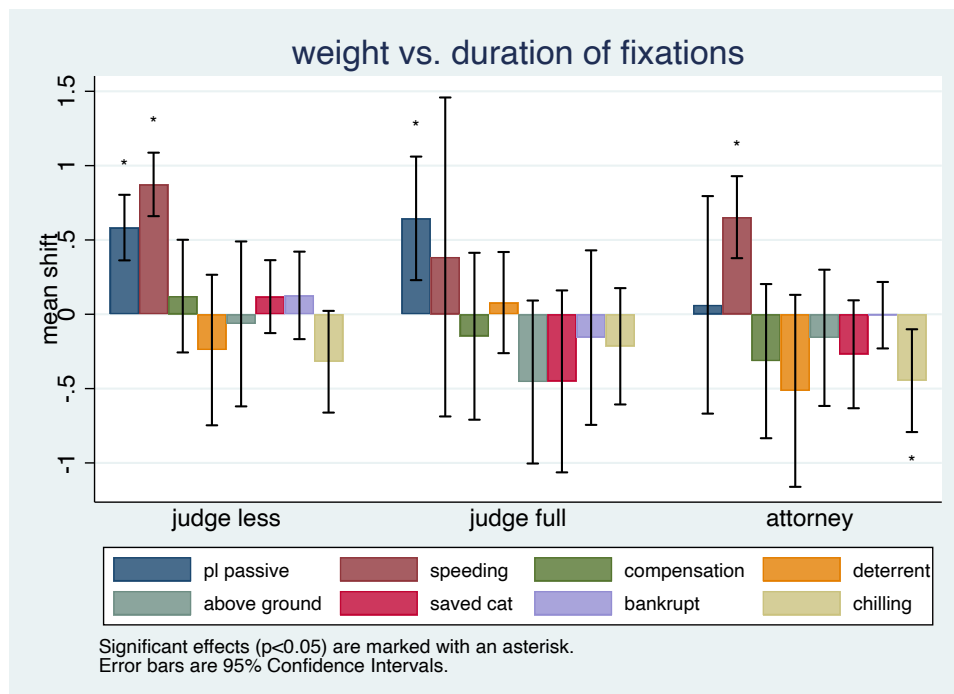


Figure 8

*Ex Post Evaluation vs. Duration of Fixations*

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€)

This explanation follows the logic of parallel constraint satisfaction. The problem is made tractable by exploiting the ambiguity. A consistent representation is found by gradually altering the impact of each item. 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. We conclude

**Result 5:** In the process of finding their decisions, participants do not elaborate systematically more on normative items speaking in favour of their decision.

### c) Response Time

If participants find the decision in the spirit of Savage, in  $H_{1c}$  we had expected that overcoming the tension between the initial evaluation of the items and the assigned role of attorney would make the task more difficult. If participants, by contrast, exploit the ambiguity by way of motivated reasoning, we had expected that knowing the desired outcome right from the start would make the problem easier. In  $H_{2c}$  we had therefore predicted that response time would be shorter for participants assigned to be attorneys. As Figure 9 shows, the distributions of response time across roles and rulings is virtually identical, except for a few outliers. If we regress response time on role and ruling, we do not find evidence of a significant difference between the response time took for judges who awarded full damages ( $\beta = -1.53$ ,  $t = 0.17$ ,  $p = 0.87$ ) or attorneys ( $\beta = 0.30$ ,  $t = 0.05$ ,  $p = 0.96$ ) compared to judges who awarded partial damages. This is further support for parallel constraint satisfaction, and in line with ( $H_{3c}$  and)  $H_{4c}$ : for a mental process relying on parallel constraint satisfaction, the exogenously imposed outcome isn't but one further constraint, taken into account when striving for a consistent representation of the inputs.

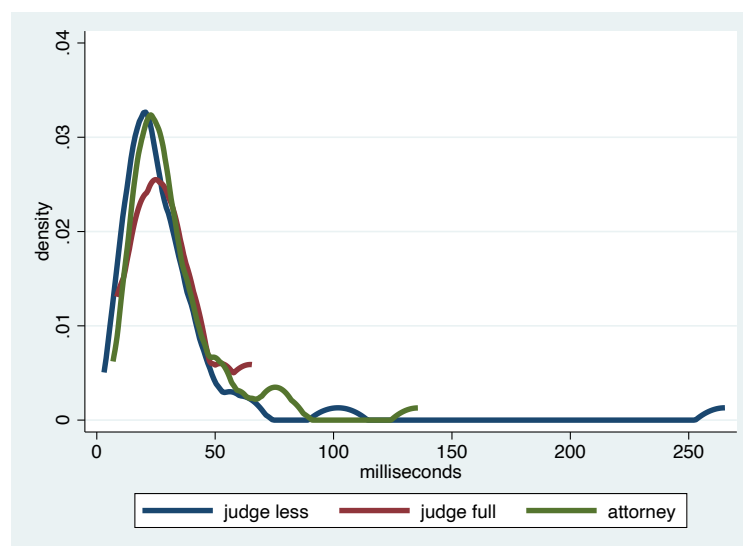


Figure 9  
Response Time

## 5. Discussion

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 full compensation is not obviously the appropriate decision.

We discuss four alternative mental paths to the solution in the face of normative ambiguity. In the logic of subjective expected utility, the decision-maker would start by assigning a normative weight to each of the competing concerns. She would apply a pre-defined meta rule to aggregate over these weighted concerns. This approach inspired by rational choice theory competes with three options relying on psychological concepts. The theory of motivated reasoning would expect that the decision-maker exploits the ambiguity to support the outcome that is most in line with her individual well-being, or her predilections. The theory of dissonance reduction expects that the decision-maker finds a way to represent the inputs such that most of the ambiguity is removed. Both theories assume an ex post correction. This is where the theory of parallel constraint satisfaction diverges. It posits that the re-evaluation of the inputs is not only a mental technique for preserving self-image and social image. The theory argues that the gradual transformation of the inputs is how the decision-maker makes the problem tractable in the first place.

We find clear evidence for parallel constraint satisfaction. Participants evaluate the normative items differently after having made a decision. This speaks against the implicit calculation of expected utility. We observe these coherence shifts for judges and attorneys. Judges have no reason to tilt the balance. This speaks against motivated reasoning. Ex post evaluations and the number or the duration of fixations are largely uncorrelated. This speaks against the mere reduction of cognitive dissonance. The number and the duration of fixations predict choices almost as precisely as the explicit reasons participants give. This shows that eye gaze is a reliable window into mental process.

Note that we find both: parallel constraint satisfaction (in the gaze data) and coherence shifts (in the re-evaluation of the inputs). Our data shows that these are distinct mental processes. Parallel constraint satisfaction is how participants have made the normatively ambiguous case tractable. Coherence shifts are how participants represent the outcome: implicitly to themselves, to maintain the self-image of being a reliable and unbiased decision-maker, and later explicitly to their assigned audiences.

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 the 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, to preserve self-image and social esteem.

Concern should focus on more subtle effects. What happens within the boundaries of plausibility? The most interesting, and the normatively most important follow-up question concerns the interaction between explicit inputs from the legal system and constraint satisfaction. One earlier finding provides a cautious note of optimism (Glöckner and Engel 2013). The design of that experiment was less involved than the present. There was no eye tracking. Only choices and coherence shifts were observed. But the experiment showed that, in a factually ambiguous case, manipulating the standard of proof (preponderance of the evidence versus beyond a reasonable doubt) had the normatively desired effect. Future research will have to systematically vary inputs from the legal system. Does it matter whether the normative expectation is spelt out in a statute or in precedent? Is it important how well the normative expectation has been translated into legal doctrine and judicial practice? Can the normative expectation be enunciated ad hoc (for instance through a change in the jurisprudence of a superior court, or through a recent ordinance)? Related to this set of follow-up questions: in which ways does legal training in the law schools affect the arguably general human ability to construct a tractable decision problem by way of constraint satisfaction?

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 ALEXANDER (1987). "Constitutional law in the Age of Balancing." Yale Law Journal **96**(5): 943-1005.
- ALEX, ROBERT (2003). "Constitutional Rights, Balancing, and Rationality." Ratio Juris **16**(2): 131-140.
- ARONSON, ELLIOT (1968). Dissonance Theory. Progress and Problems. Theories of Cognitive Consistency. A Sourcebook. R. P. Abelson, E. Aronson, W. J. McGuire et al. Chicago, Rand: 5-27.
- AWAD, EDMOND, SOHAN DSOUZA, RICHARD KIM, JONATHAN SCHULZ, JOSEPH HENRICH, AZIM SHARIFF, JEAN-FRANÇOIS BONNEFON AND IYAD RAHWAN (2018). "The Moral Machine Experiment." Nature **563**(7729): 59-64.
- BABCOCK, LINDA AND GEORGE LOEWENSTEIN (1997). "Explaining Bargaining Impasse. The Role of Self-Serving Biases." Journal of Economic Perspectives **11**: 109-126.
- BAYES, ROBIN, JAMES N DRUCKMAN, AVERY GOODS AND DANIEL C MOLDEN (2020). "When and How Different Motives Can Drive Motivated political Reasoning." Political Psychology **41**(5): 1031-1052.
- BEN-SHAHAR, OMRI AND ARIEL PORAT (2016). "Personalizing Negligence Law." New York University Law Review **91**: 627-687.
- BOHLEN, FRANCIS H (1907). "Contributory Negligence." Harvard Law Review **21**: 233-260.
- BOLSEN, TOBY AND RISA PALM (2019). Motivated Reasoning and Political Decision Making. Oxford Research Encyclopedia of Politics.
- BREHM, JACK W (1956). "Postdecision Changes in the Desirability of Alternatives." Journal of Abnormal and Social Psychology **52**(3): 384-389.
- CIALDINI, ROBERT B, MELANIE R TROST AND JASON T NEWSOM (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, JOEL AND RUSSELL H FAZIO (1984). "A New Look at Dissonance Theory." Advances in experimental social psychology **17**: 229-266.
- DE FINETTI, BRUNO (1937). "La prévision: Ses lois logiques, ses sources subjectives." Annals de l'Institute Henri Poincaré **7**: 1-68.
- DE GEEST, GERRIT (2012). "Who Should be Immune from Tort Liability?" Journal of Legal Studies **41**(2): 291-319.

- DE MOT, JEF (2013). "Comparative versus Contributory Negligence. A Comparison of the Litigation Expenditures." International Review of Law and Economics **33**: 54-61.
- DIECKMANN, NATHAN F, ROBIN GREGORY, ELLEN PETERS AND ROBERT HARTMAN (2017). "Seeing What you Want to See. How Imprecise Uncertainty Ranges Enhance Motivated Reasoning." Risk Analysis **37**(3): 471-486.
- DITTO, PETER H AND DAVID F LOPEZ (1992). "Motivated Skepticism. Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions." Journal of Personality and Social Psychology **63**(4): 568-584.
- DRUCKMAN, JAMES N AND MARY C MCGRATH (2019). "The Evidence for Motivated Reasoning in Climate Change Preference Formation." Nature Climate Change **9**(2): 111-119.
- ELLSBERG, DANIEL (1961). "Risk, Ambiguity, and the Savage Axioms." Quarterly Journal of Economics **75**: 643-669.
- ENGEL, CHRISTOPH (2007). The Psychological Case for Obliging Judges to Write Reasons. The Impact of Court Procedure on the Psychology of Judicial Decision Making. C. Engel and F. Strack. Baden-Baden, Nomos: 71-109.
- ENGEL, CHRISTOPH AND ANDREAS GLÖCKNER (2013). "Role Induced Bias in Court. An Experiment." Journal of Behavioral Decision Making **26**: 272-284.
- ENGEL, CHRISTOPH, SINIKA TIMME AND ANDREAS GLÖCKNER (2020). "Coherence-Based Reasoning and Order Effects in Legal Judgments." Psychology, Public Policy, and Law **\*\*\***: \*\*\*.
- EPLEY, NICHOLAS AND THOMAS GILOVICH (2016). "The Mechanics of Motivated Reasoning." Journal of Economic Perspectives **30**(3): 133-140.
- FELDMAN, YUVAL (2018). The Law of Good People. Challenging States' Ability to Regulate Human Behavior. Cambridge, Cambridge University Press.
- FESTINGER, LEON (1957). A Theory of Cognitive Dissonance. Evanston, Ill., Row Peterson.
- FESTINGER, LEON AND JAMES M. CARLSMITH (1959). "Cognitive Consequences of Forced Compliance." Journal of Abnormal and Social Psychology **58**: 203-210.
- GLÖCKNER, ANDREAS, TILMANN BETSCH AND NICOLA SCHINDLER (2010). "Coherence Shifts in Probabilistic Inference Tasks." Journal of Behavioral Decision Making **23**: 439-462.
- GLÖCKNER, ANDREAS AND CHRISTOPH ENGEL (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, ANDREAS AND ANN-KATRIN HERBOLD (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.
- GLÖCKNER, ANDREAS, BENJAMIN E HILBIG AND MARC JEKEL (2014). "What is Adaptive about Adaptive Decision Making? A Parallel Constraint Satisfaction Account." Cognition **133**(3): 641-666.
- GOLDBERG, JOHN CP (2015). "Inexcusable Wrongs." California Law Review **103**: 467-512.
- GREENE, JOSHUA D, FIERY A CUSHMAN, LISA E STEWART, KELLY LOWENBERG, LEIGH E NYSTROM AND JONATHAN D COHEN (2009). "Pushing Moral Buttons. The Interaction between Personal Force and Intention in Moral Judgment." Cognition **111**(3): 364-371.
- GRIESBECK, MICHAEL (2019). Die materielle Polizeipflicht des Zustandsstörers und die Kostentragungspflicht nach unmittelbarer Ausführung und Ersatzvornahmedargestellt am Beispiel der Altlasten-Problematik, Duncker & Humblot.
- HARMON-JONES, EDDIE AND JUDSON MILLS (2019). An Introduction to Cognitive Dissonance Theory and an Overview of Current Perspectives on the Theory. Cognitive Dissonance. Reexamining a Pivotal Theory in Psychology. E. Harmon-Jones. Washington, American Psychological Association: 3-24.
- HASTIE, REID (1993). Inside the Juror. The Psychology of Juror Decision Making. Cambridge England ; New York, NY, Cambridge University Press.
- HASTIE, REID, STEVEN PENROD AND NANCY PENNINGTON (1983). Inside the Jury. Cambridge, Mass., Harvard University Press.
- HOLYOAK, KEITH J. AND DAN SIMON (1999). "Bidirectional Reasoning in Decision Making by Constraint Satisfaction." Journal of Experimental Psychology: General **128**: 1-29.
- HUETTIG, FALK, CHRISTIAN NL OLIVERS AND ROBERT J HARTSUIKER (2011). "Looking, Language, and Memory. Bridging Research from the Visual World and Visual Search Paradigms." Acta Psychologica **137**(2): 138-150.
- JAMES JR, FLEMING (1952). "Contributory Negligence." Yale Law Journal **62**: 691-735.
- JUST, MARCEL A AND PATRICIA A CARPENTER (1980). "A Theory of Reading. From Eye Fixations to Comprehension." Psychological Review **87**(4): 329-354.
- KAHNEMAN, DANIEL AND AMOS TVERSKY (1979). "Prospect Theory. An Analysis of Decision under Risk." Econometrica **47**: 263-291.
- KONOW, JAMES (2005). "Blind Spots: The Effects of Information and Stakes on Fairness Bias and Dispersion." Social Justice Research **18**: 349-390.

- KÖSZEGI, BOTOND AND MATTHEW RABIN (2006). "A Model of Reference-Dependent Preferences." Quarterly Journal of Economics **121**: 1133-1165.
- KUNDA, ZIVA (1990). "The Case for Motivated Reasoning." Psychological Bulletin **108**: 480-498.
- KUNDA, ZIVA AND PAUL THAGARD (1996). "Forming Impressions from Stereotypes, Traits, and Behaviors. A Parallel-constraint-satisfaction Theory." Psychological Review **103**(2): 284-308.
- LANDES, WILLIAM M AND RICHARD A POSNER (1987). The Economic Structure of Tort Law, Harvard University Press.
- LEE, DOUGLAS G AND KEITH J HOLYOAK (2021). "Coherence Shifts in Attribute Evaluations." Decision **8**: 257-276.
- LIU, ZHUANG (2018). "Does Reason Writing Reduce Decision Bias? Experimental Evidence from Judges in China." Journal of Legal Studies **47**(1): 83-118.
- LORD, CHARLES G, LEE ROSS AND MARK R LEPPER (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, NIKLAS (1966). Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung. Berlin, Duncker & Humblot.
- MCGRATH, APRIL (2017). "Dealing with Dissonance. A Review of Cognitive Dissonance Reduction." Social and Personality Psychology Compass **11**(12): e12362.
- MOSTAFA, HESHAM, LORENZ K MÜLLER AND GIACOMO INDIVERI (2015). "An Event-based Architecture for Solving Constraint Satisfaction Problems." Nature Communications **6**(1): 1-10.
- OBERDIEK, JOHN (2008). "Philosophical Issues in Tort Law." Philosophy Compass **3**(4): 734-748.
- OBERDIEK, JOHN (2014). Philosophical Foundations of the Law of Torts, Oxford University Press.
- ORQUIN, JACOB L AND SIMONE MUELLER LOOSE (2013). "Attention and Choice. A Review on Eye Movements in Decision Making." Acta Psychologica **144**(1): 190-206.
- PENNINGTON, NANCY AND REED HASTIE (1991). "A Cognitive Theory of Juror Decision Making. The Story Model." Cardozo Law Review **13**: 519-557.
- PENNINGTON, NANCY AND REED HASTIE (1993). The Story Model for Juror Decision Making. Inside the Juror. The Psychology of Juror Decision Making. R. Hastie. Cambridge, Cambridge University Press: 192-221.
- PETERSEN, NIELS (2017). Proportionality and Judicial Activism: Fundamental Rights Adjudication in Canada, Germany and South Africa, Cambridge University Press.

- PILDES, RICHARD H (1993). "Avoiding Balancing. The Role of Exclusionary Reasons in Constitutional Law." Hastings Law Journal **45**: 711-752.
- POPPER, KARL RAIMUND (1935). Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft. Wien,, J. Springer.
- RAMSEY, FRANK P. (1931). Truth and Probability. The foundations of mathematics and other logical essays. R. B. Braithwaite. London, Routledge: 156-198.
- RAYNER, KEITH (1998). "Eye Movements in Reading and Information Processing. 20 Years of Research." Psychological Bulletin **124**(3): 372-422.
- ROSSI, FRANCESCA, PETER VAN BEEK AND TOBY WALSH (2006). Handbook of Constraint Programming, Elsevier.
- SALVUCCI, DARIO D AND JOSEPH H GOLDBERG (2000). "Identifying Fixations and Saccades in Eye-tracking Protocols." Proceedings of the 2000 Symposium on Eye Tracking Research & Applications: 71-78.
- SAVAGE, LEONARD J. (1954). The Foundations of Statistics. New York,, Wiley.
- SCHAUER, FREDERICK (1995). "Giving Reasons." Stanford Law Review **47**: 633-659.
- SCHWARTZ, GARY (1994). "Reality in the Economics of Tort Law. Does Tort Law Really Deter?" UCLA Law Review **42**: 377-444.
- SCHWARTZ, GARY T (1996). "Mixed Theories of Tort Law. Affirming Both Deterrence and Corrective Justice." Texas Law Review **75**: 1801-1834.
- SHULTZ, THOMAS R AND MARK R LEPPER (1996). "Cognitive Dissonance Reduction as Constraint Satisfaction." Psychological Review **103**(2): 219-240.
- SIMON, DAN (1998). "A Psychological Model of Judicial Decision Making." Rutgers Law Journal **30**: 1-142.
- SIMON, DAN (2004). "A Third View of the Black Box. Cognitive Coherence in Legal Decision Making." University of Chicago Law Review **71**: 511-586.
- SIMON, DAN AND KEITH J. HOLYOAK (2002). "Structural Dynamics of Cognition. From Consistency Theories to Constraint Satisfaction." Journal of Personality and Social Psychology **6**: 283-294.
- SIMON, DAN, DANIEL C. KRAWCZYK AND KEITH J. HOLYOAK (2004). "Construction of Preferences by Constraint Satisfaction." Psychological Science **15**: 331-336.

- SIMON, DAN, LIEN B. PHAM, QUANG A. LE AND KEITH J. HOLYOAK (2001). "The Emergence of Coherence Over the Course of Decision Making." Journal of Experimental Psychology: Learning, Memory and Cognition **27**: 1250-1260.
- SPAMANN, HOLGER (2020). "Extension: Lawyers' Role-Induced Bias Arises Fast and Persists despite Intervention." The Journal of Legal Studies **49**(2): 467-485.
- STULTS, DANIEL M AND LAWRENCE A. MESSE (1985). "Behavioral Consistency: The Impact of Public Versus Private Statements of Intentions." Journal of Social Psychology **125**: 277-278.
- TABER, CHARLES S AND MILTON LODGE (2006). "Motivated Skepticism in the Evaluation of Political Beliefs." American Journal of Political Science **50**(3): 755-769.
- THAGARD, PAUL AND KARSTEN VERBEURGT (1998). "Coherence as Constraint Satisfaction." Cognitive Science **22**(1): 1-24.
- THOMSON, JUDITH JARVIS (1976). "Killing, Letting Die, and the Trolley Problem." Monist **59**(2): 204-217.
- TVERSKY, AMOS AND DANIEL KAHNEMAN (1992). "Advances in Prospect Theory. Cumulative Representation of Uncertainty." Journal of Risk and Uncertainty **5**: 297-323.
- UNDERWOOD, BILL AND BERT S. MOORE (1981). "Sources of Behavioral Consistency." Journal of Personality and Social Psychology **40**: 780-785.
- YOKOO, MAKOTO AND KATSUTOSHI HIRAYAMA (2000). "Algorithms for Distributed Constraint Satisfaction. A Review." Autonomous Agents and Multi-Agent Systems **3**(2): 185-207.
- ZAMIR, EYAL AND BARAK MEDINA (2011). Law, Economics, and Morality. Oxford, Oxford University Press.
- ZIMAN, J. M. (2000). Real Science. What it Is, and What it Means. Cambridge ; New York, Cambridge University Press.

## Appendix

### Preregistered Hypotheses

Evaluation: 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.

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.

Decision Time: Participants who have been assigned to represent plaintiff decide faster.

### Supplementary Data Analysis

	pl passive	speeding	compensation	deterrent	above ground	saved cat	bankrupt	chilling	N
<b>number of fixations</b>									
judge less	.471*** (.129)	.709*** (.134)	.167 (.154)	-.141 (.154)	-.027 (.161)	.094 (.127)	-.069 (.198)	-.404* (.179)	46
judge full	.516* (.230)	.542+ (.301)	.071 (.213)	.036 (.207)	-.289* (.123)	-.396+ (.225)	-.068 (.206)	-.392+ (.231)	14
attorney	-.128 (.242)	.496** (.161)	-.159 (.194)	-.387 (.276)	-.146 (.180)	-.182 (.173)	-.021 (.107)	-.456** (.159)	51
<b>duration of fixations</b>									
judge less	.583*** (.108)	.874*** (.105)	.122 (.186)	-.241 (.249)	-.064 (.272)	.119 (.120)	.127 (.145)	-.319+ (.168)	46
judge full	.645*** (.185)	.385 (.479)	-.148 (.250)	.079 (.152)	-.456+ (.273)	-.452+ (.273)	-.157 (.262)	-.216 (.175)	14
attorney	.063 (.361)	.653*** (.136)	-.315 (.256)	-.515 (.318)	-.159 (.226)	-.269 (.179)	-.006 (.110)	-.447** (.170)	51

Table 3

*Ex Post Evaluation of Normative Weight vs. Eye Gaze*

linear structural models (multivariate regressions) with only a constant for each of the dependent variables testing whether the difference is significantly different from zero

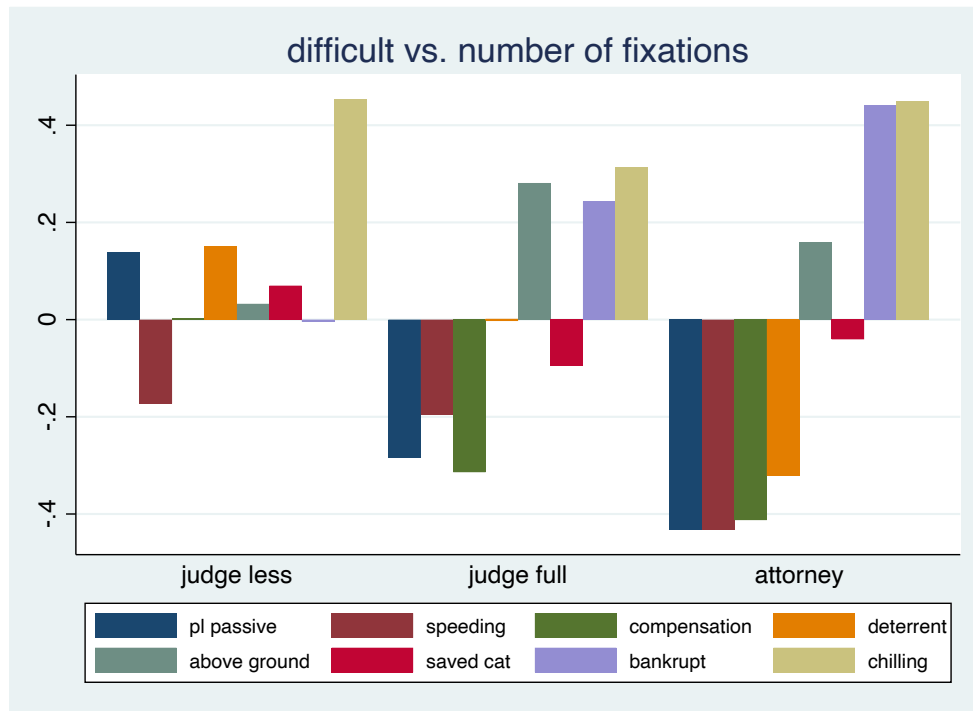


Figure 10

*Ex Post Evaluation of Difficulty vs. Number of Fixations*

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€)

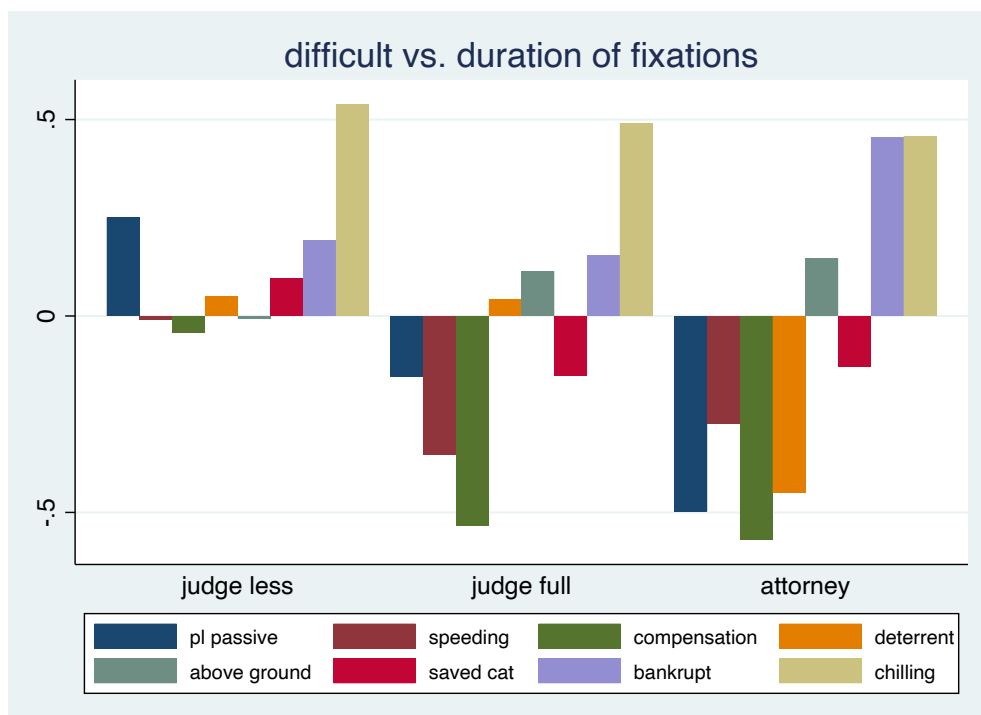


Figure 11

*Ex Post Evaluation of Difficulty vs. Duration of Fixations*

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€)

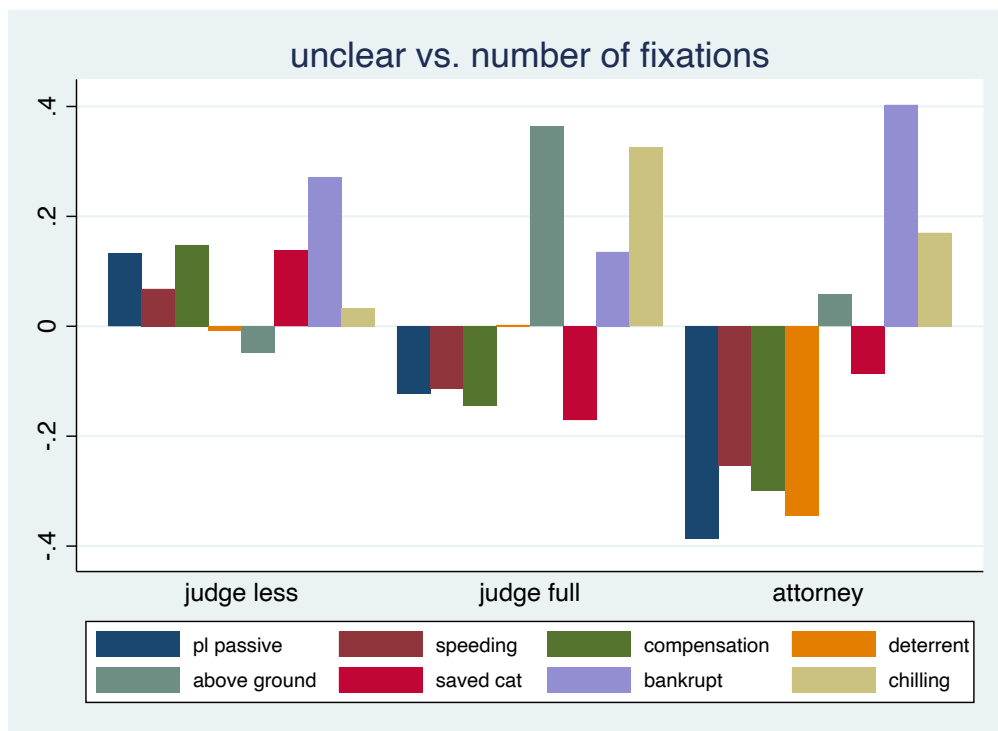


Figure 12

*Ex Post Evaluation of Lack of Clarity vs. Number of Fixations*

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€)

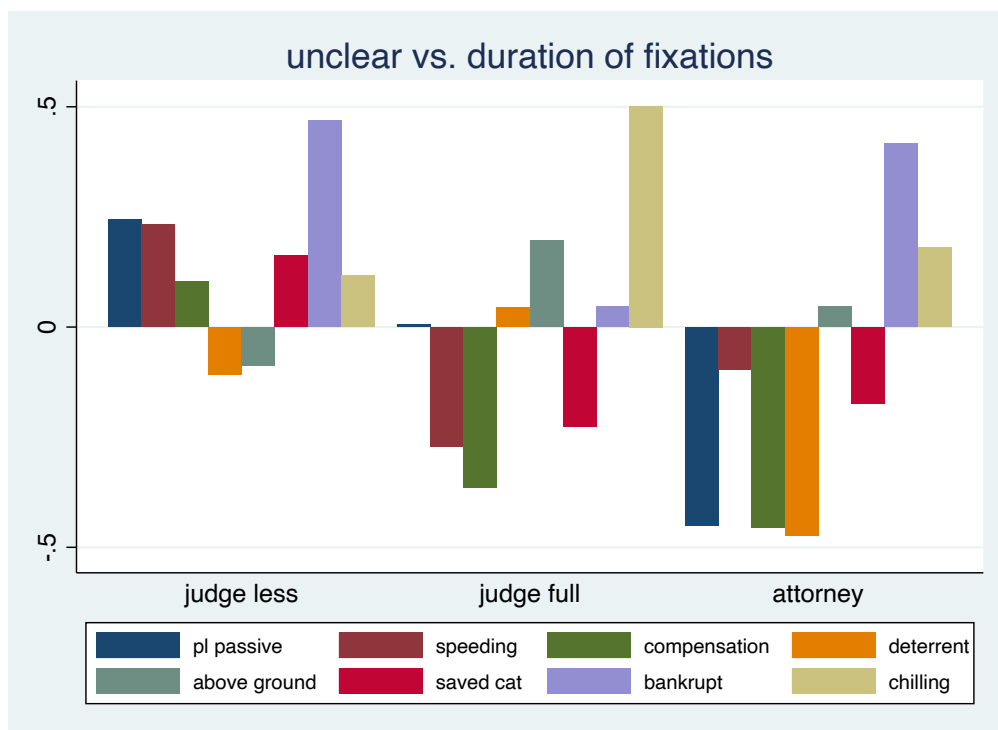


Figure 13

*Ex Post Evaluation of Lack of Clarity vs. Duration of Fixations*

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€)

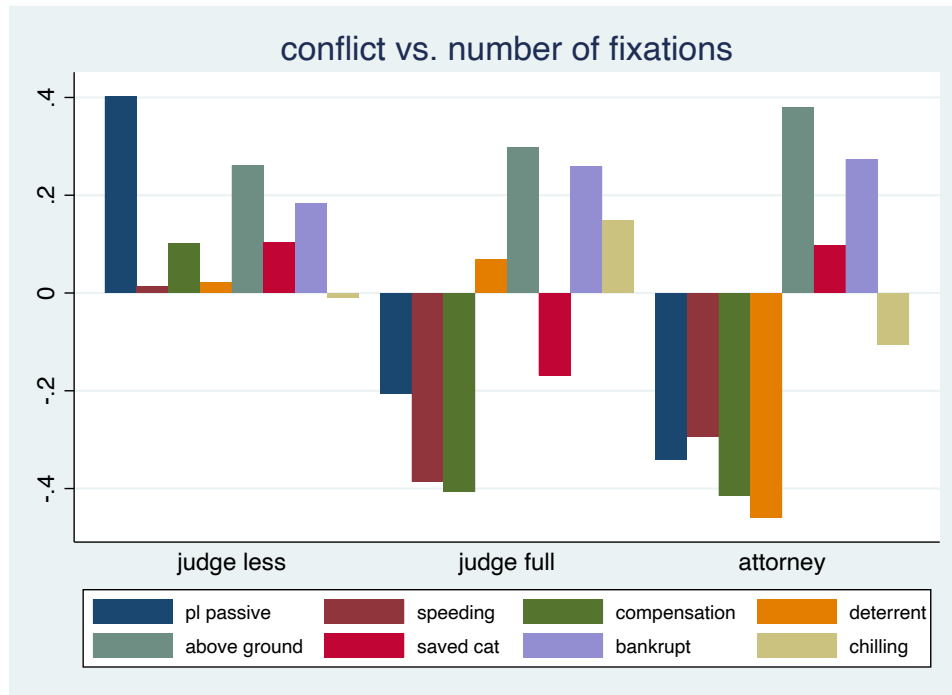


Figure 14

*Ex Post Evaluation of Degree of Conflict with Other Items vs. Number of Fixations*

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€)

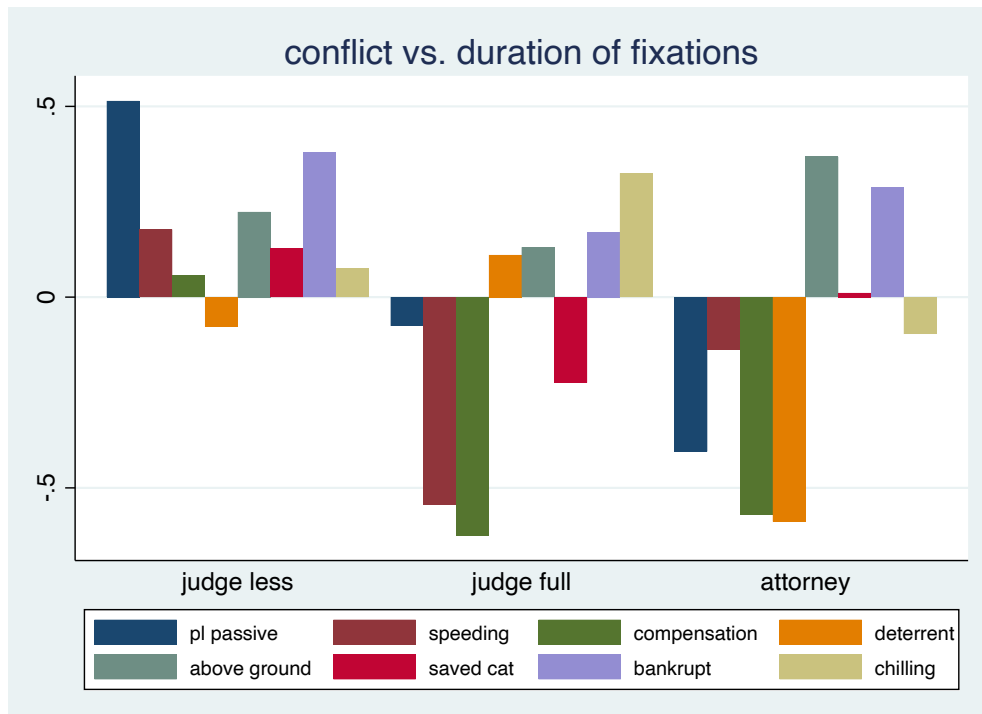


Figure 15

*Ex Post Evaluation of Degree of Conflict with Other Items vs. Duration of Fixations*

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€)