Order Effects and the Evaluation Bias in Legal Decision Making

Bartosz W. Wojciechowski[1], Lee C. White[2], Carsten Allefeld[2], & Emmanuel M. Pothos[2]

1. Faculty of Management and Social Communication, Institute of Applied Psychology, Jagiellonian University, Kraków, Poland.

2. Department of Psychology, City, University of London, London, UK.

*Correspondence: Emmanuel Pothos, emmanuel.pothos.1@city.ac.uk

Word count: 8181

Abstract

There has been intense interest in biases in legal decision making, such as order effects and evaluation biases (biases arising from making judgments, as opposed to just observing some information). We extend previous work in three ways. First, we employ a population sample including judges, prosecutors, and attorneys, as well as naïve participants, to investigate the extent of biases for legal professionals. Second, we use realistic materials, summaries of real legal cases. Finally, we study two biases, order effects and the Evaluation Bias, the latter being a bias corresponding to more extreme evaluations if a previous, oppositely valenced piece of information had been evaluated vs. just observed. Both biases were reliably observed across all groups of legal professionals and a group of lay participants; there was no evidence that different groups of participants displayed either of the two biases to a lesser extent. The presence of two, basic decision biases in a study involving realistic legal stimuli and with legal professionals raises questions about the robustness of decision processes in the legal system.

1. Introduction

1.1 General problem

The underlying assumption of the judiciary is that all judgments rendered by a court, as well as the interpretation, application, and enforcement of existing law, are all effects of rational decision making. In the course of a case, a judge or magistrate can make several important decisions, sometimes quickly and under pressure (Dhami, 2003). For example, a judge decides if an accused can stay out of jail pending trial, whether or not evidence is admissible, and if evidence should be included in a case. Judges, as well as administrators who establish the relevant rules and processes, have complex roles. Part of their responsibility is to think as applied cognitive psychologists, by recognizing the limitations of mental operations (Saks & Spellman, 2016).

There has been extensive research on biases in legal decision making. There has been work on external factors, such as pre-trial publicity, racial stereotypes, and eyewitness credibility based on spurious characteristics (Spellman & Tenney, 2010). Other work has revolved around memory problems, such as lawyers having difficulty judging whether a witness correctly identifies an alleged perpetrator and whether a confession offers a truthful account of a suspect's deeds (Simon, 2012; cf. Howe, 2013). Research in simulated and real legal settings has revealed biases in court rulings (Green & Wrightsman, 2003), including cases of judges explaining a defendant's actions based on their own past experiences (Saks & Thompson, 2003). Lawyers sometimes underutilize probabilistic information and fail to understand relevant statistical principles (McAuliff, Nementh, Bornstein, & Penrod, 2003; Daftary-Kapur, Dumas, & Penrod, 2010; Spellman & Tenney, 2010). Overall, even with identical information, differing verdicts may be reached (Green & Wrightsman, 2003).

The extent of research regarding biases in legal decision making is bewildering; is more research needed? We think yes, to better match experimental procedures to courtroom situations and concerning novel ideas, especially if they bring together apparently independent results.

We consider order effects and evaluation biases. Order effects concern biases from the order in which pieces of information are presented. Evaluation biases concern whether one piece of information is evaluated (e.g., with a preliminary decision), as opposed to just observed, before another piece of information is considered. A practical reason for studying these biases is that they are ways to manipulate legal proceedings. Attorneys and prosecutors might manipulate evidence order, to increase impact with the jurors. For example, rules of conduct and evidence discovery assume that, when filing an indictment, the prosecutor should present all the evidence at that stage. However, sometimes evidence is initially withheld and revealed at later stages, as part of strategy.

Indeed, the legal systems of most countries do not impose sanctions for selective disclosure of evidence. Also, preparing a defense includes a strategy for presenting evidence. Analogously, regarding evaluation biases, attorneys or prosecutors might attempt to introduce questions e.g. regarding guilt vs. innocence, as legal proceedings unfold, with a view to bias subsequent evidence assessment.

1.2 Biases from order and evaluations

McKenzie et al. (2002; Trueblood and Busemeyer, 2011) asked participants to imagine they were jurors in a criminal case and to rate confidence in guilt, following the presentation of evidence in different orders. They observed an order effect, in that the probability of guilt given information from strong prosecution followed by weak defense was 0.72, but the same information presented in the converse order inflated the probability of guilt to 0.75, which is a weak recency effect (Anderson, 1959; Wilson, 1971). Dahl et al. (2009) and Price and Dahl (2013) also reported a recency bias: the perceived credibility of two types of evidence—one incriminating and the other exonerating—depended on the order in which they were presented, with the evaluated evidence being viewed as more credible if it was presented last rather than first. Note, in this work it was assumed that an order effect would emerge from contrast, when there is a contradiction between the two pieces of evidence (cf. Scherer & Lambert, 2009). Finally, in Maegherman et al. (2022), the likelihood of guilt increased when incriminating evidence was presented last and decreased when exonerating evidence was presented last.

Recency biases could arise from recollection ease. Costabile and Klein (2005) showed that incriminating evidence was more likely to result in a guilty conviction, if introduced late in the trial as opposed to early on. Regardless of whether the evidence was found to be admissible or not, this pattern persisted. Subsequent analysis revealed that the jurors' recollections of important evidence may have contributed to this result.

Contrary to the above conclusions, Marksteiner et al. (2011) and Pennington (1982) argued for primacy effects. Procedure and methodological variation across research may account for a portion of these discrepancies; in some cases the emergence of primacy vs. recency effects has been tied to theoretical accounts. Marksteiner et al. (2011) offered the idea of asymmetric processing and asymmetric skepticism (Ask & Granhag, 2007; Ask et al., 2008). They observed that an initial hypothesis that a suspect is guilty resulted in rating incriminating evidence as more reliable, than exonerating evidence.

Pennington and Hastie (1992) suggested that jurors' decisions are influenced by explanations—semantic frameworks concerning the causal linkages among the events the decision maker believes occurred. Witness credibility, assessments of confidence, and perceptions of evidence strength are all affected by the ease of tale construction. Arguing against universal primacy or recency effects, Pennington and Hastie (1992) reported that when the evidence is arranged according to a narrative, individuals reach stronger, more confident conclusions in line with the preponderance of the evidence. They argued that their results cannot be explained in terms of different pieces of evidence being more memorable than others. For example, they demonstrated how using a story construction method enhanced the impact of the information "completing" the story and how offering direct story inferences influenced decisions toward the more comprehensible story. Pennington and Hastie (1992) concluded that the best "order of proof" during a trial is a narrative story sequence.

Relatedly, Charman et al. (2015, 2016) suggested that order effects arise as context effects, in that context can alter how one evaluates a piece of evidence or how different pieces of data are integrated together. In Charman et al. (2015), DNA evidence indicating guilt had more significant influence on a subsequent alibi evaluation, than exonerating DNA evidence. That is, evaluation of an alibi was impacted by participants' knowledge of a previous piece of evidence, e.g., the result of DNA testing (see also Kassin et al., 2013).

Regarding order effects, in Charman et al.'s (2015, 2016) work, the valence of the evidence and the sequence in which it was presented both acted as moderators regarding the overall impact of context. The same evidence led to different decisions, depending on presentation order. Relatedly, according to Ask et al. (2011), context effects are typically the consequence of superficial processing and only manifest when the important evidence comes after, not before, the evidence that supports a suspicion of guilt. Furthermore, context effects are noticed when the underlying view is one of guilt because guilt beliefs elicit stronger biases in the evaluation of future data than beliefs of innocence.

Lagnado and Harvey (2008) examined the interaction between context and order effects, in relation how people revise their beliefs when evidence is discredited. Mock jurors read simplified criminal cases and judged the probability that a suspect was guilty from sequentially presented evidence. Regardless of whether the first piece of information related to a later discredited item, discrediting the later item reduced belief in the first item. This extension effect depended on how temporally close the original and discrediting evidence appeared and whether their valence was consistent (e.g., all pointing towards guilt). The latter finding indicates that aligned pieces of information (e.g., consistently pointing towards guilt) cohere together, regardless of causal relations

between them. These ideas point to an explanation of order effects in terms of the interaction between stage of processing and grouping of pieces of evidence together.

Are there reports of lack of order effects? Medical and legal decision making are similar in that in both there is an expectation of rationality. Bergus et al. (1998; Bergus et al., 2002) reported a recency order effect in a medical diagnosis task, with medical professionals ("family physicians"). Heard, Rakow, and Foulsham (2018) re-examined the original paradigm, using formats of presentation ostensibly better aligned with reading biases. Heard et al. (2018) found no order effects, a result partly explained in terms of such presentation differences.

Overall, there is considerable evidence for order effects in legal decision making. There is less work concerning evaluation biases, that is biases arising from whether some piece of information is used in an evaluation (e.g., a preliminary evaluation of guilt) vs. just observed. Charman et al. (2016) suggested that participants would analyse pieces of evidence fairly thoroughly, unless there is a prior conviction about a suspect's guilt, in which case evidence consideration is superficial. Though Charman et al. (2016) did not make this suggestion, we could argue that a prior belief in guilt could arise from an early judgment. Pennington and Hastie (1992) reported that (mock) jurors making a final overall judgment were more likely to follow an explanation-based judgment strategy and that a "wait until the end" strategy increased confidence, compared to a cumulative, item-by-item updating judgment strategy. Carlson and Russo (2001) showed that the interpretation of identical pieces of evidence can depend on whether participants had already formed a tentative verdict or not: having a tentative verdict would bias interpretation towards consistency with that verdict. Heard et al. (2018) suggested that one of the reasons why they did not replicate Bergus et al.'s (2002) recency effect is that Bergus et al. (2002) asked participants for an initial judgment, which might have served as a starting anchor.

A step-by-step processing strategy means that a preliminary judgment is generated after each piece of evidence, while with an end-of-sequence one there is a single final judgment. Hogarth and Einhorn (1992) considered several studies presenting information sequentially and for which it was possible to make a characterization as step-by-step vs. end-of-sequence. Their conclusion was that step-by-step processing appears to lead to recency (cf. Heard et al.'s, 2002, suggestion).

Yearsley and Pothos (2016) explored variants of step-by-step processing modes of presentation, regarding pieces of evidence relevant to a (hypothetical) crime. They reported that the density of intermediate judgments for guilt vs innocence reduced the probability of change from the initial belief concerning the suspect. Such an evaluation bias cannot be interpreted as recency or primacy, because it shows an independence of presentation format, not differential weighting of earlier vs. later pieces of information.

In summary, there is much evidence for order effects in legal decision making, though there are inconsistencies concerning direction and explanation: some researchers have reported recency effects (Costabile & Klein, 2005; McKenzie et al., 2002), while others have considered how earlier information biases later interpretations (Charman et al., 2015, 2016; Lagnado & Harvey, 2008; Pennington & Hastie, 1992). Sometimes, the consideration of order effects is tied to evaluation biases, e.g. because of an interaction between preliminary judgments and later perception of information (Ask et al., 2011; Carston & Russo, 2001; Hogarth & Einhorn, 1992).

1.3 Theoretical motivation

Memorability of the presented information seems a plausible, but perhaps partial explanation, for recency effects (e.g., Costabile & Klein, 2005; Price & Dahl, 2013, though note the latter suggested that recency effects could also be due to contrast, if the last and first pieces of evidence have conflicting valence). But it is also clear that there are more complex interactions between beliefs and evidence presented at different points in a sequence. Asymmetric processing (Marksteiner et al., 2011) and asymmetric skepticism (Ask & Granhag, 2007; Ask et al., 2008) are biases whereby an initial belief for e.g. guilt can lead to biased processing for corresponding evidence. Such ideas relate to coherence models of decision making (Holyoak & Simon, 1999; Glöckner, Betsch, & Schindler, 2010; Simon et al., 2001), according to which perception of information is affected by beliefs, so that information and beliefs align with each other (cf. cognitive dissonance, Festinger, 1957). Also related is Thagard's (2005) default pathway of accepting assertions, when the source is reliable and the assertions are consistent with our beliefs. One common thread here is that if we are committed to a belief, then we process the available information in a way that reduces tension between beliefs and perceptions. There have been several expressions of this idea in legal decision making, such as from Carston and Russo (2001) and Lagnado and Harvey (2008).

An alternative way to explain order effects is with narratives (Pennington & Hastie, 1992): the weighting of different pieces of information is partly determined by fits with the overall narrative of previous evidence. Drawing from work in social psychology more generally, it seems plausible that, in a sequence, each piece of information activates unique thoughts, which influence the processing of subsequent information (Schwarz, 2007). Therefore, different sequences can result in different narratives (Ask et al., 2011; Charman et al., 2015, 2016).

Can order effects be explained from the relative salience of different pieces of information? For example, cues higher in utility would be used more frequently in decision problems (Newell et al., 2004) and the correlation between cues and outcomes seems to provide the best account of how participants utilise cues in an inference problem (Rakow et al., 2005), at least in the context of the

specific tasks in the corresponding studies (for these two studies, changes in share price of fictional companies). However, it is hard to see how such ideas can result in order effects, unless cue evaluation is itself subject to context.

An important theoretical challenge is when can we conclude that legal decision making is rational, which is especially important in legal or medical decision making. Most researchers consider Bayesian probability theory as the appropriate rational standard (Griffiths et al., 2010; Oaksford & Chater, 2007), a position which can be justified in multiple ways, including evolutionary considerations (Ramirez & Marshall, 2017). Order effects challenge Bayesian expectation (Lagnado & Harvey, 2008; Pennington & Hastie, 1992). For two pieces of evidence on whether a suspect is guilty, Bayesian theory requires that $Prob(Guilt|evidence1, evidence\ 2) = Prob(Guilt|evidence2, evidence\ 1)$; this is a trivial implication of commutativity in conjunction in Bayesian theory. One could write $Prob(Guilt|evidence1, evidence\ 2, order1) \neq Prob(Guilt|evidence2, evidence\ 1, order2)$, so as to create order effects by explicitly conditionalizing on order. However, such an approach is post hoc as there is no guidance for when to expect recency, primacy, or no order effects.

There is some work arguing for the use of heuristics in legal decision making, which challenge expectation for consistency with Bayesian (rational) principles. Dhami (2003) investigated decisions in UK courts, concerning the imposition of punitive bails. There were two candidate explanations, Franklin's rule and a matching heuristic. The former is rigid and requires the compensatory combination of several differentially weighted cues. The latter involves selection between a subset of cues with predictions based on one cue (Gigerenzer et al., 1999). Dhami (2003) reported support for the matching heuristic. It is unsurprising that legal decision making is subject to biases (cf. Newell et al., 2004). For example, Gigerenzer's work has made the case for 'fast and frugal' heuristics, especially under cognitive load – Dhami (2003) noted that the briefness of the bail hearings and the urgency with which decisions had to be made contributed to biasing judicial rulings.

An interesting aspect of Dhami's (2003) work is that biases were demonstrated for magistrates, for whom we would have high expectations for rational decision making. There is related evidence. Helm et al. (2016) reported that 'elite' arbitrators, specializing in resolving business disputes, would rely too much on intuition and decision fallacies, just like judges. As Guthrie et al. (2001) argued, especially judges are more driven to reach better conclusions and have more time to do so (as well as more support from clerks), nevertheless their decisions are often driven by intuition (Rachlinski & Wistrich, 2017) and are characterized by several decision biases (Guthrie et al. considered five). As Frank (1949, p. 410) noted some 70 years ago: "When all is said and done, we

must face the fact that judges are human." Dhami (2003) concluded that magistrates' actions went against the principles of due process, which hold that the number of innocent defendants subjected to harsh punishment should be kept to a minimum.

It is possible to understand order effects using a probability theory alternative to Bayesian theory, quantum theory -- the probability rules from quantum mechanics, without any of the physics. In quantum theory some questions are so-called incompatible and some are compatible, while in Bayesian theory all questions are compatible. Incompatible questions cannot be resolved concurrently; responding to one, introduces uncertainty for the other. Researchers working with quantum models have claimed that quantum theory is a way to formalise ideas like the ones from Schwarz (2007) or Festinger (1957; for overviews see Busemeyer & Bruza, 2011; Pothos & Busemeyer, 2022). Explanations about order effects from quantum models essentially assume that incompatibility drives 'interference' between pieces of evidence presented in particular orders. But note that consistency with quantum models does not imply rationality, without additional assumptions (Pothos et al., 2017).

The situation regarding evaluation biases is similar. Some of the explanations for order effects apply here too. For example, a (preliminary) decision that the suspect is guilty potentially creates new thoughts or perspectives, affecting later ones (Carston & Russo, 2001; Lagnado & Harvey, 2008; see also Holyoak & Simon, 1999; Glöckner, Betsch, & Schindler, 2010; Simon et al., 2001).

As for order effects, we can ask whether evaluation biases are rational: in Bayesian theory, measurements (e.g., a decisions) do not have a functional role. For example, in a legal case, a juror might believe that the suspect is guilty with 65% probability. If at that point they are asked for a binary judgment, they might toss a suitably weighted mental coin. Nonetheless, there is no requirement from Bayesian theory for a change in the underlying belief state. Bayesian theory is not inconsistent with changes, e.g., $Prob(guilt) >< Prob(guilt|decision)$, but the theory offers no guide for how such changes would occur.

In a quantum model, resolving a question requires that the state identifies with the question outcome. For example, think of a person in an art gallery looking at a painting. Their mental state will reflect some uncertainty about whether they like the painting or not. If the person is asked whether they like the painting, on responding e.g. with a yes, the mental state changes to identify with liking (this is Luder's projection postulate in quantum theory). Therefore, in standard quantum theory there is a requirement that the mental state changes in a particular way, as a result of measurements.

One question is whether the particular evaluation biases reported in legal decision making (Carston & Russo, 2001; Pennington & Hastie, 1992; Charman et al., 2016) can be explained using quantum theory. There is some equivalence between quantum theory and other accounts, in that an early judgment of e.g. guilt could impact on the perception of later information. Indeed, these investigators have invariably mentioned anchoring effects, whereby an early judgment of guilt implies differential weighing of condemning vs. exonerating evidence. However, existing ideas have been invariably expressed without formal models -- formal models (whether quantum or Bayesian) are not necessarily at odds with such ideas, but it is difficult to be more specific (Pothos & Busemeyer, 2022).

White, Pothos, and Busemeyer (2014) produced a specific evaluation bias prediction from quantum theory, for pairs of stimuli of opposite valence; in their original experiments, valence was positive vs. negative affect, but other judgments have been explored (White et al., 2015; 2020). We refer to this prediction as the Evaluation Bias (capitalized so as to distinguish it from more general evaluation biases). The Evaluation Bias is that in a pair of oppositely valenced stimuli such that the second one is always evaluated, if the first one is evaluated too (as opposed to just observed), then the judgment for the second one is more extreme. For example, if the valence of the second stimulus is negative, then more extreme would mean that it is even more negative, than without the intermediate rating.

Figure 1 is a caricature of how quantum theory could apply to the Evaluation Bias and illustrates some of the main ideas. There are three main elements in the diagram. First, there is a representation of the mental state, denoted as ψ (and variants, e.g., ψ_g). Second, we have question outcomes, corresponding to one-dimensional subspaces (also called rays). The question outcomes in Figure 1 are whether a suspect is guilty or innocent. The probability of different question outcomes depends on the *overlap* between the mental state and the corresponding question outcome (probability is computed as the squared length of the *projection* of the mental state vector onto a ray). Third, the impact of evidence introducing evidence towards e.g. guilt corresponds to a rotation of the mental state towards, in this case, the guilt ray. Let us consider a legal case composed of two parts, a part indicating guilt followed by one indicating innocence. We first consider the first part and so our mental state is set close to the Guilty ray (ψ_g). Then, we receive the second part and the mental state rotates to $\psi_g{}'$, towards the Innocent ray. After the second part, we are asked for a judgment, whose strength for innocence corresponds to red length along the Innocent ray. If after the first part we are asked for an intermediate judgment, it is most likely that we will respond guilty, so that the mental state becomes a normalized vector along the Guilty ray. Then, we have the same rotation towards the Innocent ray (cf. Stewart, Brown, & Chater, 2005), which means that the new

mental state is $\psi_g{}''$. This mental state is closer to the Innocent ray and so the corresponding projection is more consistent with innocence (the blue length). Overall, the intermediate judgment creates a more intense evaluation for the second one. There are many additional details here which are needed for a model of an Evaluation Bias (see White et al., 2020). Note, to account for order effects, similar representations and processes from quantum theory can be used (e.g., Trueblood & Busemeyer, 2011).

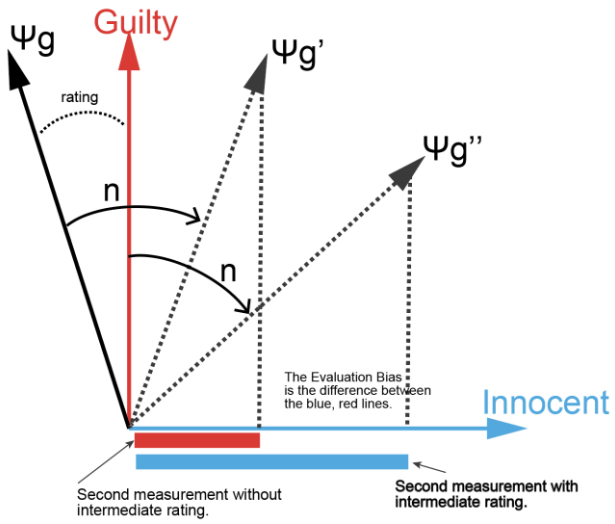

Figure 1. An example of how quantum theory is broadly consistent with the Evaluation Bias. The letter 'n' denotes the rotation towards the Innocent ray, as a result of receiving the second part of a legal case (in this example, this second part is assumed to indicate innocence).

1.4 Extending previous work

We believe there are three priorities concerning future work. First, there is the issue of rationality. There is no escaping the fact that non-Bayesian reasoning is not rational, according to well-established definitions of rationality. Perhaps, for example, fast and frugal heuristics (Gigerenzer et al., 1999) are good enough for everyday decision making, but in a court of law, with decisions impacting people's lives, it is reasonable to expect a high standard of rationality. Heuristics are likely to impact on the accuracy of criminal verdicts and undermine the adjudicative process, so that it falls short of the precision expected in the criminal justice system (Simon, 2012).

However, most legal decision making studies have been conducted with naïve participants, usually university undergraduates, who are asked to pretend they are mock jurors. This does have some validity: in common law countries, when a jury is present, the main decisions are made by members of a jury and a judge only advises and guides jurors. Accordingly, the focus on non-

professionals in legal decision making, conducted in the UK, USA, and Canada is reasonable, since they are (sometimes) the actual decision makers. Nevertheless, there is also interest in whether legal professionals are prone to decision biases, especially given that biases could be employed (e.g., by attorneys) for particular effects. Of course, the difficulty with a sample of legal professionals is accessibility (Dhami, 2003). In the present work, we address this challenge.

Second, much legal decision making work has involved artificial scenarios of, arguably, low importance and interest to participants. Forensic psychologists have questioned whether a reliable examination of legal decision making is possible, regardless of the characteristics of the experimental materials and participants (Konecni & Ebbesen, 1979; Kapardis, 2003). For example, psychological research conducted solely with student samples pretending to be mock jurors and exposed to greatly simplified materials may mischaracterize behavior compared to what we would expect in real court cases (Daftary-Kapur, Dumas, & Penrod, 2010; Fox, Wingrove, & Pfeifer, 2011; Spellman & Tenney, 2010). In this work, we adopted materials based on real criminal legal cases.

Third, we attempt a preliminary link of some biases together, employing insights from quantum theory. Yearsley and Trueblood (2018; see also Wojciechowski & Pothos, 2018) demonstrated that conjunction fallacies were correlated with order effects, in a decision making task related to the US Presidential elections. The basis of this prediction was that, according to quantum theory, both effects arise from incompatible representations for the questions. Analogously, regarding order effects and the Evaluation Bias, according to quantum theory a common cause of both effects is quantum-like representations and processes. Note, the mechanism which leads to order effects is different from that responsible for evaluation biases/ the Evaluation Bias (interference vs. collapse). There are some a priori reasons why quantum theory might be relevant in legal decision making, if such decision making suffers from information overload and pressured, rushed conditions (Trueblood et al., 2017; Pothos et al., 2021; Yearsley and Trueblood, 2018).

2. Experimental design

2.1 Participants

Four groups of participants took part in the experiment: 40 criminal court judges (22 women, 18 men, aged between 29 and 66 years, M=42; SD=7.7; with professional experience from 3 to 30 years, M=14; SD=6.3); 18 prosecutors (7 women, 11 men, aged between 29 and 60 years, M=40; SD=8.37 with professional experience raging from two to 35 years, M=10; SD=8.21) and 22 attorneys (9 women, 13 men; aged between 28 and 58 years, M=38; SD=6.15; with professional experience raging from three to 34 years, M=11; SD=7.35). Finally, we recruited 40 participants without legal

background (21 women, 19 men, aged between 21 and 62 years, M=32; SD=11.1; their non-legal professional experience ranged from none to 40 years; M=10; SD=10.8). All participants were recruited in Poland.

The first author (an attorney at law since 2011) used his professional experience and network to identify participants (judges, prosecutors, and defense attorneys). This was essential, because complications arise in any payment to a judge or a prosecutor and, indeed, it seems unlikely that a judge could be incentivized to take part in a psychology study for a small payment. Since it was not possible to pay some of the participants, we decided not to pay any of the participants. The participants with no legal background were recruited by the first author amongst colleagues/ acquaintances. Participation was voluntary, there was no compensation. Prospective participants were informed that the aim of the research was to explore legal decision making and the distinctive characteristics of judges and lawyers in evidence evaluation.

Ethics approval for this work was provided by the Ethics Committee of the Institute of Psychology at the Silesian University in Katowice, Poland and all participants provided written consent prior to participation.

The sample size was based on practical considerations: we recruited the largest number of legal professionals we had access to. Note, sample sizes are analogous to those in previous work with the Evaluation Bias (White et al., 2014), though in this previous work the design was fully within participants, whereas presently it is mixed effects.

Finally, we note that this and the other experiments in this work were not pre-registered.

2.2 Design

The present experiment had a 2 (evidence order: guilty-innocent, innocent-guilty) x 2 (rating condition: single vs. double rating, to mean either one rating for both pieces of information at the end or a preliminary rating after the first piece of information followed by a final rating) x 4 (role: judge vs prosecutor vs attorney at law vs layperson) x 6 (case; there were six cases) mixed effects design (Figure 2). All factors were between participants. There was a single random effect corresponding to the individual participants.

Note that the experimental design had two separate and independent parts, leading to two different datasets. The first dataset was analyzed and considered in Wojciechowski and Pothos (2018). The present report concerns the second dataset, related to tests of the Evaluation Bias and order effects. The reason why two separate experiments were run in this way is because of the difficulty of recruiting participants in the legal profession.

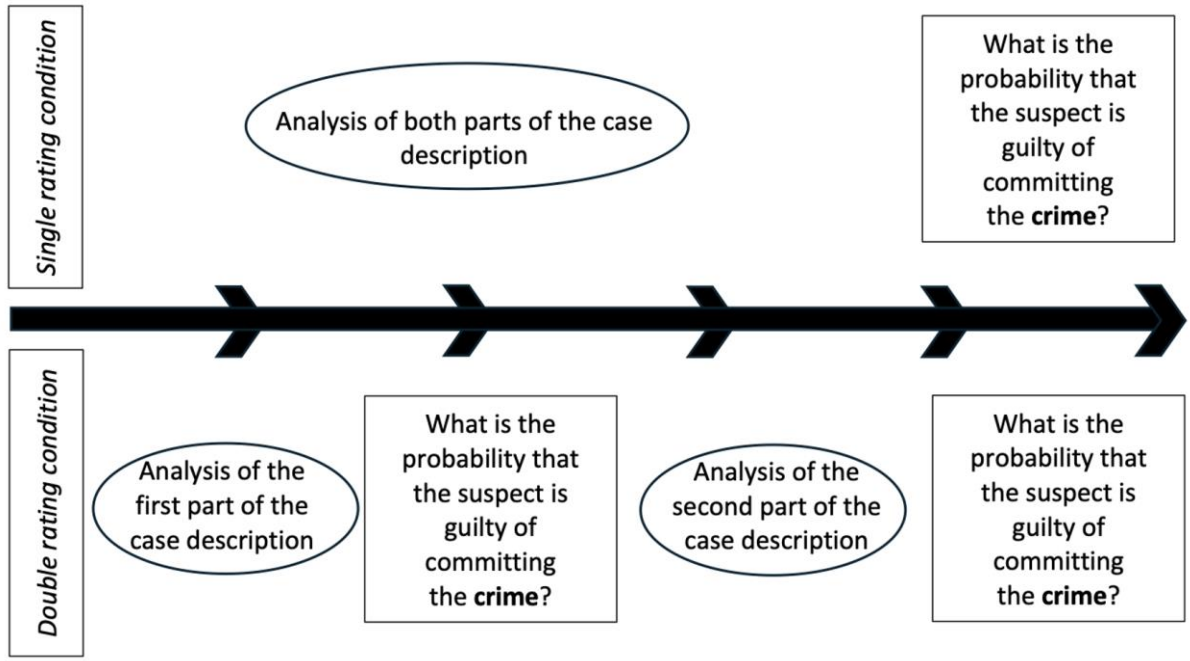

Figure 2. The design of the study and an outline of the procedure.

2.3 Materials

As part of a previous project (Wojciechowski & Pothos, 2018), we were granted access to real criminal case files from various District Courts and Appellate Courts in Poland, pertaining to crimes committed between 2000 and 2015 (see Appendix A for some notes on the legal context). Wojciechowski and Pothos (2018) describe in full how the materials for this work were selected and a summary is provided below. We gathered all pertinent information for each criminal case, including interrogation and interview protocols, expert testimony, and adjudications. Then we picked 40 cases from 400 possibilities, based on the following criteria. First, the case had to be from a different court district than the one the legal professional participants were from. Second, we aimed for approximately equal proportions of guilty and innocent defendants, based on the information included in the justifications of the judgments (this is part of the court's ruling). Third, we verified that accurate suspect assertions were supported by the available evidence and false statements were refuted by the evidence. We made no attempt to balance the gender, age, and ethnicity of the suspects for the 40 selected case transcripts.

We then assessed whether corresponding case summaries correctly revealed the innocence or guilt of the suspect. This was established by having two independent, competent raters examine each summary and determine whether it led to a conclusion of suspect's guilt or innocence. The two raters were an experienced retired prosecutor and an experienced retired attorney at law specializing in criminal cases, both with over 30 years of experience in the Polish justice system. Only

criminal case summaries of confirmed valence during this preliminary stage (of selecting the 40 case transcripts), for which Kendall's concordance coefficient was over .75, were used in the main experiment (note, this is just a non-parametric statistic for rank correlation, commonly used for assessing agreement amongst raters and inter-rater reliability). Six cases were finally selected, three for which the first part of the description was incriminating and the second and half of the description was exonerating, and three with reverse order (all participants rated the same cases, but the main factors – evidence order and rating condition – had to be manipulated between participants). All original materials were in Polish (translations in Supplementary Electronic Material 1).

2.4 Procedure

Following agreement to take part in the study, participants received the materials (the same six criminal case summaries for all participants) printed out on paper, in booklets delivered in sealed envelopes. They were told they could go through the case materials when and where this suited them best, the common assumption being that they would do so at home. After completing the tasks, after around two weeks, the experimenter collected the booklets in sealed envelopes to assure participants (specifically judges, prosecutors, and attorneys) that any ratings will not be linked to them. We felt this was the only possible way to engage busy legal professionals and, therefore, we had no control over the location and time of the assessment.

Participants were expected to read all materials and, based on the available evidence, rate the guilt of the suspects on a 1 to 10 scale, with 1 corresponding to definitely innocent and 10 to definitely guilty. Based on participant reports, we estimated that participants took between 20 and 50 minutes to read the summaries and evaluate the cases.

The evidence for each suspect was organized into two parts, which had to be read sequentially. Regarding rating condition, in all six cases, participants were asked to provide an overall rating, after having read both parts. Each participant, for three of their cases, had to rate the first part as well (double rating condition) and for the remaining three of their cases only provided a final, overall rating (single rating condition). Regarding evidence order, each participant saw three cases conforming to innocence-guilt (IG) and three to guilt-innocence (GI) (there were three more 'filler' cases in a GG order). That is, both rating condition and evidence order were manipulated between participants. This was undesirable, but inevitable. In the procedure of White et al. (2014, 2020), the design was fully within participants, that is, the same participant would rate two stimuli in both the double and the single ratings condition (pairs of stimuli would be shown twice, interspersed

by many other pairs). However, in the present case, this was not possible, because of the distinctiveness of the stimuli.

A limitation in the procedure is that it is not possible to independently verify that participants processed the two pieces of information for each case in the intended order. However, there are three mitigating considerations. First, participants were volunteers and it could be assumed that they would make some effort to be cooperative and follow the instructions. Second, while some participants might violate the instructions (e.g., process the two parts in an order different to the one intended and/or not make the intermediate ratings), wholesale inconsistency with the instructions would make it impossible to observe either an order effect or an Evaluation Bias. Finally, we discuss in Section 2.6 a follow-up, in-lab study, which replicates the order effect and the Evaluation Bias, in the first testing phase (in this in-lab study, there were two testing phases, attempting to collect more measurements per participant for each case).

2.5 Results

The data for all experiments is available from the authors. The dependent variable was the second (overall) rating of guilt for each of the six cases, which corresponded to a number between 1 and 10. We applied a linear transformation to the [0,1] range, simply for stylistic reasons, so that the dependent variable resembles a probability – since the transformation is linear, statistical conclusions are identical between the original and transformed variables. There were four independent variables, rating condition, evidence order, role (judge vs prosecutor vs attorney at law vs layperson), and case (six levels). We considered a single random effect, participants.

We first identified the best linear model for assessing the four fixed effects in our design, using a best model selection procedure, based on comparing nested models with -2 log likelihood and the chi squared statistic (Appendix B). The best model included all fixed effects and all interactions, with the random effect of participants modeled with random intercepts only. All F-tests in this and subsequent sections are from this model. Note, we follow standard practice in mixed effects analyses in not reporting effect sizes. This is because the goal with effect sizes, in general, is to relate a measure of effect (e.g., a regression coefficient) to a measure of random variation, and in a mixed effects model there are several sources of random variation. Put differently, standardized effect sizes remove the influence of the sample size, but in mixed effects models there are several sample sizes, e.g. trials per participant and number of participants (Jiang & Nguyen, 2021; Peugh, 2010).

Regarding role, of primary interest are the three-way interaction with rating and order (do participants in the four different groups display the Evaluation Bias to a varying degree?) and the two-way interaction with order (do participants in the four different groups display order effects to a varying degree?). We consider these results in the subsections for the Evaluation Bias and order effects. Here, we note that there was a just significant effect of role, $F(3, 110)=2.636$, $p=.05$. There was a trend for attorneys to offer less guilty verdicts, compared to participants in the other three groups (the means for the other three groups were nearly identical; for attorneys, prosecutors, judges, and lay persons and means were respectively .46, .51, .50, .50). There was also a main effect of case and an interaction between case and role, respectively $F(5,589)=40.677$, $p<.001$ and $F(15,587)=2.221$, $p=.005$, indicating that some cases attracted more guilty ratings than others and that groups varied concerning which cases were considered more guilty.

As a methods check, we can ask whether when the first part of a case was G there was a higher guilty rating for that first part, than when the first part of a case was I. This was the case, with the differences in means being M=.76, SD=.24 vs M=.36, SD=.25; a corresponding mixed effects pairwise comparison (fixed effect of valence, random effect of participants, no slopes) was significant, $F(1, 352)=231.4$, $p<.001$.

2.5.1 Evaluation Bias

One aim is to examine if the Evaluation Bias can be observed in legal decision making. The Evaluation Bias concerns whether there is a difference in overall judgements about guilt, depending on whether or not the participant rated the first piece of information. In the GI condition we expect mean ratings in the double rating condition to be lower than those in the single rating condition. In the IG condition we expect the mean ratings in the double rating condition to be greater than those in the single rating condition. These predictions are an implication of the rating scale we employed, so that judgments of guilt corresponded to numerically higher ratings (cf. White et al., 2020). Therefore, the test for the Evaluation Bias concerns the interaction between order and rating condition, which was significant, $F(1,620)=8.526$, $p=.004$. Figure 3 shows a pattern of results consistent with the Evaluation Bias. The three-way interaction between order, rating condition, and role was non-significant, $F(3,620)=2.343$, $p=.072$, showing no evidence that different groups of participants showed the Evaluation Bias to a different degree. However, the four-way interaction with order, rating condition, role, and case was significant, $F(14,620)=1.835$, $p=.031$. It appears that for different cases the extent of the Evaluation Bias for stronger for some groups, than for others.

Overall results Results by role

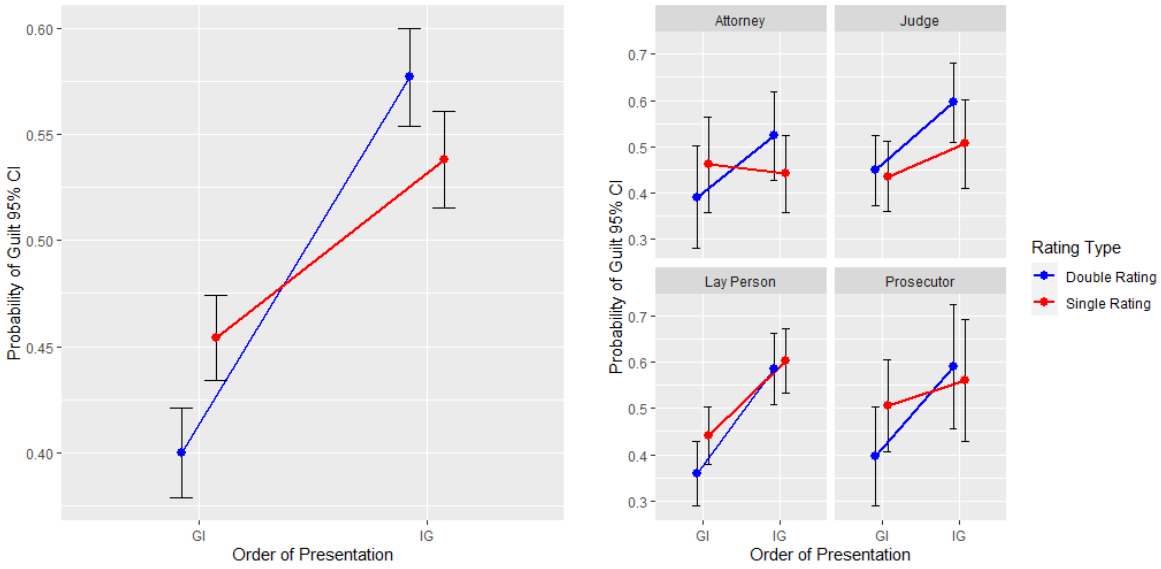

Figure 3. Overall and by role mean participant final ratings of single and double rated GI and IG stimuli.

2.5.2. Order effects - asymmetries in processing information

Do perceptions of guilt depend on whether the initial information points towards guilt vs. innocence? The test for the order effect concerns the main effect of order, that is, whether participants considered the two parts for each case in the IG vs. GI order. Recall, in all cases, participants were instructed to offer an overall evaluation of guilt, by taking into account both pieces of information. Therefore, if there are no order effects, there should be equality in the overall/ second ratings, across the two presentation orders, IG and GI. Conversely, for example, consider the GI condition: if the first judgment (based on guilty information) influences to a greater extent the second judgment (which indicates innocence), then the second judgment would be higher (indicating more guilt) than the second judgment in the IG condition (for the same case). As shown in Figure 4, there was a clear recency effect, that is the valence of the last piece of information has a higher weight in the overall evaluation, compared to the first piece of information, $F(1, 620)=54.397$, $p<.001$.

The order effect did not vary by role, that is, the two-way interaction between order and role was not significant, $F(3,620)=2.334$, $p=.073$. However, the three-way interaction between order, role, and case was significant, $F(15,620)=2.472$, $p=.002$, that is, different groups of participants displayed order effects more strongly for some cases, than for others.

Overall results Results by role

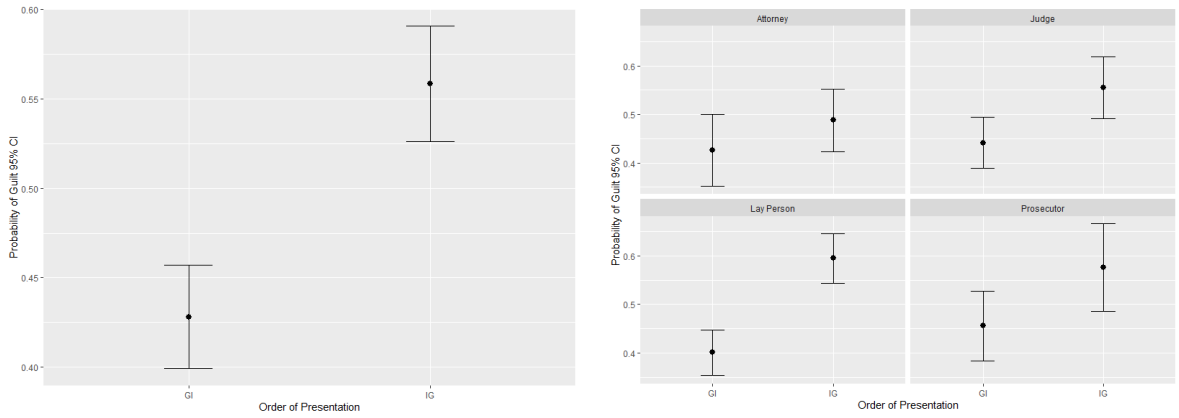

Figure 4. By order and by role ratings for GI and IG stimuli. Ratings were combined across single and double rating condition.

2.5.3. Evaluation Bias and order effects – are they related?

This analysis is preliminary. We are interested in whether there is a relationship between the Evaluation Bias and order effects, that is, whether the presence of one effect makes more likely the other effect. Recall, each participant saw six cases, so that each case was viewed in one of two presentation orders, with responses consisting either of a single overall rating at the end or, in addition, a preliminary rating after the first piece of evidence. Accordingly, it is not possible to investigate a putative relationship between order effects and the Evaluation Bias, within participants. Instead, we adopted an item-based analysis, averaging responses across participants, to compute measures of order effects and Evaluation Bias, for each of the six cases: the specific hypothesis is that, *if* certain cases encourage more quantum-like processing, then for these cases there should be both higher order effect and a higher Evaluation Bias, and vice versa. What makes the analysis preliminary is that it is based on only six data points.

Regarding a measure of order effects, in the main analysis we computed order effects across both rating conditions (both single and double). We focus here on just the double rating condition, because there is a higher order effect in this condition and, given the very small N, we need all the sensitivity we can get. We subtracted GI overall ratings from IG second ratings (regardless of condition). Note, computing order effects in the IG-GI order is appropriate, since quantum theory is more consistent with recency effects and the rating scale was set up so that guilty judgments corresponded to higher ratings. If there were no order effects, this measure would be zero. Regarding the Evaluation Bias measure, recall that the Evaluation Bias is that the double rating condition produces more extreme ratings, relative to the single rating one. Therefore, we can

compute an Evaluation Bias measure as differences concerning the overall (final) rating for each item, as follows:

in the GI condition, rating in the single rating condition minus rating in the double rating one

in the IG condition, rating in the double rating condition minus rating in the single rating one

Since for each item we wanted a single measure of Evaluation Bias strength, we averaged the above two differences.

The correlation between the order effect and Evaluation Bias measure was r=0.84, p=.03 (given the small sample size, the significance value is offered as tentative). Looking at this result as linear regression (Figure 5), for every unit of increase in the Evaluation Bias variable, the predicted order effect variable increased by about 2.20 units; the Evaluation Bias explained a fairly large portion of variance in the order effect variable (R^2=0.71) and the effect size is fairly large (Cohen's f^2=2.44). Of course, the very low sample size raises concerns about the robustness of this result, which we cannot circumvent in this paradigm. We employed bootstrapping analyses (1,000 replications) as a robustness check for the slope and intercept estimates, which produced regression estimates similar to what we had before (β_0=0.10, β_1=2.20). The bootstrap-derived mean estimate for the slope was 2.198, with a bias of -0.150, indicating a slight underestimation, and a standard error of 1.363 (bias, in this context, refers to the difference between the average value of the bootstrap estimates and the original regression estimate). Similarly, the intercept had a bootstrap estimate of 0.100, with a bias of 0.003 (indicating a minimal overestimation) and a standard error of 0.047. The results add support to the stability of the regression estimates. Overall, this result can only be offered as preliminary. An additional reason highlighting the preliminary nature of this result is that if we compute order effect across both rating conditions, the order effect is weaker and the correlation with Evaluation Bias is lower (and not significant).

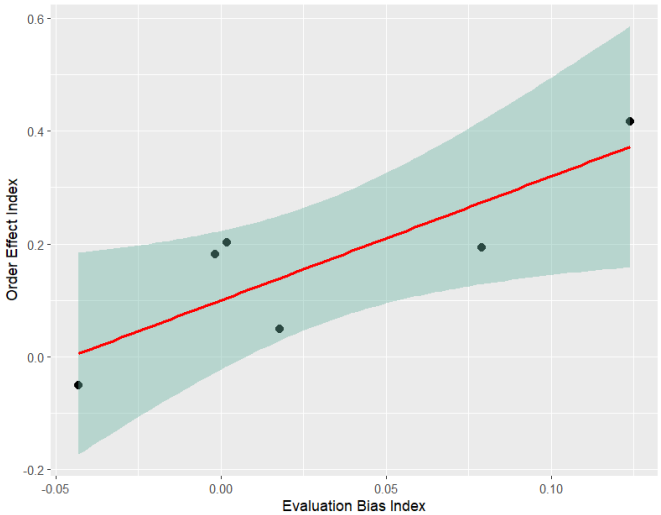

Figure 5. Results for correlation of Evaluation Bias index and order effect index. The shaded band indicates a 95% confidence interval for the correlation.

2.6 Pilot experiment: can we recast the study with a within participants design?

Two main limitations of our experimental work (rigorous control regarding the procedure and the putative relatedness between the Evaluation Bias and order effects, see Sections 2.4 and 2.5.3 respectively) could be addressed if the experimental design were fully within participants, so that the same participants judged legal cases in both orders (GI, IG) and under both rating conditions (single, double). Therefore, instead of having each participant receive each legal case once (with a random assignment to the order and rating conditions), they would receive each case four times, (two levels of order times two levels of rating condition). Following an editorial comment, we partially examined this possibility, by having each participant respond to each case twice. This pilot is reported fully in Appendix C; here, we summarise the main points.

The pilot had two presentation stages. In the first presentation stage, the materials and procedure were identical to those of the main study (Sections 2.3, 2.4). The second presentation was also nearly identical to the first: the six legal cases participants received would have some minor, superficial differences (e.g., different names), to reduce recollection of the earlier case; we randomly varied the order and rating condition for each case, so that, for example, if one participant responded to a particular case in the IG order and with a single rating, in the second presentation one of these factors would randomly be flipped – it is in this way that the design in this pilot was more within participants, compared to the main study. As another way to reduce recollection, the two presentation stages were three weeks apart. All participants in the pilot were psychology students. Testing was carried out with an experimenter present, under controlled conditions.

To summarise the results, in the first presentation stage, we replicated both the order effect and the Evaluation Bias, but in the second presentation case only the former (see Appendix C for details). Moreover, and despite our efforts, many participants commented that they recognised the legal cases from the first presentation stage. While it is encouraging that the two effects do replicate, it unfortunately remains hard to see how we can extend the main study so that participants respond to each case multiple times. The problem is that the use of realistic legal materials (which we think is an important step towards higher ecological validity) means that each case is highly memorable (contrast with White et al., 2014, 2020).

3. Discussion

Despite the wealth of evidence regarding biases in legal decision making, we offered three motivations for the present work. First, much research on legal decision making has been conducted with non-legal professionals (there are exceptions, e.g., Rachlinski & Wistrich, 2017), so there is a need to expand the knowledge base accordingly. Second, legal professionals engaging with realistic materials might offer reasoning better approximating real legal decision making (Fox et al., 2011). Finally, we can ask whether we can link some biases together and approached this challenge in an exploratory way using quantum theory (Pothos & Busemeyer, 2022).

We recruited a sample of attorneys, prosecutors, judges, and legally lay people and utilized materials which were summaries of previous legal cases. Judges and lay participants would have the least vested interest in the outcome, while prosecutors and attorneys might be biased towards particular outcomes and so perhaps more likely to employ influencing strategies, including intermediate evaluations and different orders (Dahl et al., 2009; Price & Dahl, 2013; Maegherman, et al., 2021; Devine et al., 2001; Lawson, 1968). Judges represent the ultimate expectation concerning bias-free decision making. There might be similar expectations about rationality for attorneys and prosecutors, but their role in the dispute is different – the role of the court is to search for the ground truth, but the role of attorneys/ prosecutors is to defend/ prosecute and so such individuals may be selective towards information which favours their standpoint.

Our results revealed evidence both for an Evaluation Bias and a recency order effect, but no interactions with participant role. Observing an Evaluation Bias in legal decision making extends our understanding of the circumstances which can give rise to such biases (White et al., 2020). As for order effects, while there have been plenty of previous studies showing order effects, our results reinforce these previous conclusions with more ecologically valid sampling and materials (see also Enescu & Kuhn, 2012). Interestingly, we identified interactions with role and case for both the Evaluation Bias and order effects, showing that different groups of participants displayed these biases to a greater or lesser extent, for different cases. These differences might be due to variations in professional and personal experience, as well as the litigation roles typically performed. The cases were selected with a mindset to make the Evaluation Bias and order effects plausible, but they varied widely (as would be expected with realistic materials) and so it is unsurprising that different participant groups approached them somewhat differently. Judges, lawyers and prosecutors have different roles to play in criminal trials. When analysing the materials, they tend to focus on those aspects of the description that involve a substantive legal assessment and determine the premises of criminal liability. To the layperson, the case summaries were plausibly morality stories, judged by personal experiences, beliefs, and values.

Regarding the possible interactions just with role, for order effects, the averaged difference in the final rating across the two different orders was 0.21 for lay participants and 0.12, 0.08, 0.14 for prosecutors, attorneys, and judges, respectively (these order effects represent differences in numbers which are in the [0,1] range). That is, the size and direction of the order effect was similar across all cases. For the Evaluation Bias, the pattern of results for judges and lay participants offers a different impression from that for attorneys and prosecutors (Figure 3). These trends merit further examination. A speculative hypothesis is that certain kinds of biases might be easier to suppress than others, but since there was no significant interaction with role, further work is needed. It is tempting to suggest that a future iteration of this study with better sampling might clarify the situation. However, the sampling limitations in the present work will be difficult to address, given the problem of recruiting legal professionals.

Despite our efforts to study order effects and the Evaluation bias in an ecologically valid way for legal decision making, there are several constraints regarding the generality of the results. Most obviously, the legal professionals were recruited as part of the professional network of the first author, which might have impacted on how they approached the study. Additionally, the case materials were simplified, lacking much of the context and information that would apply in real legal cases. Despite our attempts to simplify and standardise the materials, the cases were not as well matched as typical experimental materials. Regarding procedure, a potential limitation is that we do not have certainty that our instructions were followed as stated. However, based on our results, we think this is unlikely: the absence of Evaluation Bias or order effects could indicate evidence against such effects *or* failure to comply with the instructions. Given that we did observe both an Evaluation Bias and order effects, we cannot see how it is possible that the experimental materials were processed in a way other than the intended one. Also, we replicated the order effect and the Evaluation Bias, under controlled experimental conditions, with the same materials and lay participants (Section 2.6). Another limitation is that real legal cases would not in general reflect the IG, GI structure in the present work.

According to quantum theory decision models, quantum-like representations should be the cause of both the Evaluation Bias and order effects. We offered some preliminary evidence that this is the case. However, because each case was evaluated only once by each participant, it is not possible to compute a measure of order effects and the Evaluation Bias within participants. This precludes a rigorous test of this idea (Huang et al, 2023; Trueblood et al., 2017; Yearsley & Trueblood, 2018). The data needed to assess the relatedness between the two effects could be provided from a fully within participants version of the present experiment. However, in Section 2.6,

we presented a failed pilot showing that, without radical re-imagining of the experiment this is not possible.

Overall, despite these limitations, we hope that the demonstration of the Evaluation Bias and order effects, with legal professionals and with realistic legal stimuli, provides useful additions to the empirical literature, while the consideration of quantum theory offers some promise for a more principled understanding of some of the relevant biases.

Acknowledgments

EMP was supported by European Office of Aerospace Research and Development (EOARD) grant FA8655-23-1-7220.

References

Anderson, N. H. (1959). Test of a model for opinion change. Journal of Abnormal and Social Psychology, 59, 371–381. DOI: 10.1037/h0042539

Ask, K., & Granhag, P. (2007). Motivational bias in criminal investigators' judgments of witness reliability. Journal of Applied Social Psychology, 37, 561–591. DOI: 10.1111/j.1559-1816/2007.00175.x

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. Applied Cognitive Psychology, 22, 1245–1259. DOI: 10.1002/acp.1432

Ask, K., Granhag, P. A., & Rebelius, A. (2011). Investigators under influence: How social norms activate goal-directed processing of criminal evidence. Applied Cognitive Psychology, 25, 548–553. DOI: 10.1002/acp.1724

Basieva, I., Pothos, E. M., Trueblood, J. Khrennikov, A., & Busemeyer, J. R. (2017). Quantum probability updating from zero priors (by-passing Cromwell's rule). Journal of Mathematical Psychology, 77, 58-69. DOI: j.jmp.2016.08.005

Bergus, G. R., Levin, I. P., & Elstein, A. S. (2002). Presenting risks and benefits to patients. Journal of General Internal Medicine, 17, 612–17. DOI: j.1525-1497.2002.11001.x

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and order of information. Medical Decision Making, 18, 412–417. DOI: 10.1177/0272989X9801800409

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. Journal of Experimental Psychology: Applied, 7, 91-103. DOI: 10.1037/1076-898X.7.2.91

Charman S.D., Carbone J., Kekessie S., Villalba D.K. (2015). Evidence Evaluation and Evidence Integration in Legal Decision – Making: Order of Evidence Presentation as a Moderator of Context Effects. Applied Cognitive Psychology, 30. DOI: 10.1002/acp.3181

Charman S.D., Carbone J., Seyram K., Villalba D.K. (2016). Evidence Evaluation and Evidence Integration in Legal Decision-Making: Order of Evidence Presentation as a Moderator of Context Effects. Applied Cognitive Psychology, 30, 214 – 225. DOI: 10.1002/acp.3181

Costabile, K. A., & Klein, S. B. (2005). Finishing strong: Recency effects in juror judgments. Basic and Applied Social Psychology, 27, 47–58. DOI:10.1207/s15324834basp2701_5

Daftary-Kapur T., Dumas R., Penrod S.D. (2010). Jury decision-making biases and methods to counter them. Legal and Criminological Psychology, 15, 133-154. DOI: 10.1348/135532509X465624

Dahl, L. C., Brimacombe, C. A., & Lindsay, D. S. (2009). Investigating investigators: How presentation order influences participant-investigators' interpretations of eyewitness

identification and alibi evidence. Law and Human Behavior, 33, 368–380. DOI: 10.1007/s10979-008-9151-y

Devine D.J., Clayton L.D., Dunford B.B., Seying R. & Pryce J. (2001). Jury Decision Making. 45 Years of Empirical Research on Deliberating Groups. Psychology, Public Policy and Law, 7, 622-727. DOI: 10.1037/1076-8971.7.3.622

Dhami, M. K. (2003). Psychological models of professional decision-making. Psychological Science, 14, 175-180.

Elqayam, S., & Evans, J. S. B. T. (2013). Rationality in the new paradigm: strict versus soft Bayesian approaches. Thinking and Reasoning, 19, 453 – 470. DOI: 10.1080/13546783.2013.834268

Enescu, R., & Kuhn, A. (2012). Serial effects of evidence on legal decision making. The European Journal of Psychology Applied to Legal Context, 4, 99–118.

Festinger, L. (1957). A Theory of Cognitive Dissonance. Stanford Univ. Press, Stanford.

Fox, P., Wingrove, T., Pfeifer, C. (2011). A Comparison of Students' and Jury Panelists' Decision-making in Split Recovery Cases. Behavioral Sciences and the Law, 29, 358–375. DOI: 10.1002/bsl.968

Frank, J. (1949). Courts on Trial: Myth and Reality in American Justice. Princeton University Press, Princeton.

Frederick, S. (2005). Cognitive reflection and decision making. The Journal of Economic Perspectives, 19, 25–42. 10.1257/089533005775196732

Gigerenzer, G., Todd, P. M., the ABC Research Gropu (Eds.) (1999). Simple heuristics that make us smart. New York: Oxford University Press.

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. Journal of Behavioral Decision Making, 23, 439 – 462. DOI: 10.1002/bdm.668

Green E. & Wrightsman L. (2003). Decision Making by Juries and Judges: International Perspective, in: Handbook of Psychology in Legal Contexts, eds. Carson D., Bull R. John Wiley and Sons, Chichester, pp. 401 - 422.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364. DOI: 10.1016/j.tics.2010.05.004

Guthrie, C., Rachlinski, J.J., & Wistrich A.J. (2001). Inside the Judicial Mind. Cornell Law Faculty Publications, Paper 814; http://scholarship.law.cornell.edu/facpub/814

Hatvany, N., & Strack, F. (1980). The impact of a discredited key witness. Journal of Applied Social Psychology, 10, 490-509. DOI: 10.1111/j.1559-1816.1980.tb00728.x

Heard, C. L., Rakow, T., & Foulsham, T. (2018). Understanding the effect of information presentation order and orientation on information search and treatment evaluation. Medical Decision Making, 38, 646-657. DOI: 10.1177/0272989X18785356

Helm, R.K., Wistrich, A.J., & Rachlinski, J.J. (2016). Are Arbitrators Human? Journal of Empirical Legal Studies, 13, 666 – 692. DOI: 10.1111/jels.12129

Hertwig, R., Hoffrage, U., & the ABC Research Group (2013). Simple heuristics in a social world. New York: Oxford University Press. Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. Journal of Experimental Psychology: General, 128, 3–31.

Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating: the belief-adjustment model. Cognitive Psychology, 24, 1-55. DOI: 10.1016/0010-0285(92)90002-J

Howe, M. L. (2013). Memory development: implications for adults recalling childhood experiences in the courtroom. Nature Reviews Neuroscience, 14, 869-876. DOI: 10.1038/nrn3627

Huang, J., Busemeyer, J. R., Ebelt, Z. & Pothos, E. M. (2023). Quantum Sequential Sampler: a dynamical model for human probability reasoning and judgments. In Proceedings of the 2023 Annual Conference of the Cognitive Science Society. Sydney, Australia: Cognitive Science Society.

Jaeger Ch.B. & Trueblood J.S. (2019). Thinking Quantum: A New Perspective on Decision making in Law; Florida State University Law Review, 46, 733 – 806.

Jiang, J. & Nguyen, T. (2021). Linear and Generalized Linear Mixed Models and Their Applications, Springer.

Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Kapardis A. (2003). Psychology and Law. A Critical Introduction. Cambridge University Press, Cambridge.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. Journal of Applied Research in Memory and Cognition, 2, 42–52; DOI: 10.1016/j.jarmac.2013.01.001

Konecni V.J., Ebbesen E.B. (1979). External Validity of Research in Legal Psychology. Law and Human Behavior, 3, 39 – 70. DOI: 10.1007/BF01039148

Lagnado D.A., Harvey N. (2008). The impact of discredited evidence. Psychonomic Bulletin & Review, 15, 1166 – 1173. DOI: 10.3758/PBR.15.6.1166

Lawson G (1968). Order of Presentation as a Factor in Jury Persuasion. Kentucky Law Journal, 56, 523 - 555.

Lenth R (2023)._emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.5, <https://CRAN.R-project.org/package=emmeans>.

Maegherman E., Ask K. Horselenberg R., van Koppen P. (2022). Law and order effects: on cognitive dissonance and belief perseverance. Psychiatry, Psychology and Law, 29, 33 – 52. DOI: 10.1080/13218719.2020.1855268

Marksteiner T., Ask K., Reinhard M.A., Granhag P.A. (2011). Asymmetrical Scepticism Towards Criminal Evidence: The Role of Goal- and Belief-Consistency. Applied Cognitive Psychology, 25, 541 – 547. DOI: 10.1002/acp.1719

McAuliff B.D., Nemeth R.J., Bornstein B.H., & Penrod S.D. (2003). Juror Decision - Making in the Twenty - First Century: Confronting Science and Technology in Court, [in:] Handbook of Psychology in Legal Contexts, eds. Carson D., Bull R., John Wiley and Sons, Chichester, pp. 303 - 372.

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. Journal of Behavioral Decision Making, 15, 1–18. DOI: 10.1002/bdm.400

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: the success of "success". Journal of Behavioral Decision Making, 17, 117-137. DOI: 10.1002/bdm.465

Oaksford, M. & Chater, N. (2007). Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press.

Pennington, D. C. (1982). Witnesses and their testimony: Effects of ordering on juror verdicts. Journal of Applied Social Psychology, 12, 318–333. DOI: 10.1111/j.1559-1816.1982.tb00868.x

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. Journal of Personality and Social Psychology, 62, 189–206. https://doi.org/10.1037/0022-3514.62.2.189

Peugh, J. L. (2010). A practical guide to multilevel modeling. Journal of School Psychology, 48, 85-112. DOI: 10.1016/j.jsp.2009.09.002.

Pinheiro J, Bates D, R Core Team (2023)._nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-162, <https://CRAN.R-project.org/package=nlme>.

Pothos, E. M. & Busemeyer, J. M. (2022). Quantum cognition. Annual Review of Psychology, 73, 749-778. DOI: 10.1146/annurev-psych-033020-123501

Pothos, E. M., Busemeyer, J. R., Shiffrin, R. M., & Yearsley, J. M. (2017). The rational status of quantum cognition. Journal of Experimental Psychology: General, 146, 968-987. DOI: 10.1037/xge0000312

Pothos, E. M., Lewandowsky, S., Basieva, I., Barque-Duran, A., Tapper, K., & Khrennikov, A. (2021). Information overload for (bounded) rational agents. Proceedings of the Royal Society B, 288, 20202957. DOI: 10.1098/rspb.2020.2957

Price, H. L., & Dahl, L. C. (2013). Order and strength matter for evaluation of alibi and eyewitness evidence. Applied Cognitive Psychology, 28, 143–150. DOI: 10.1002/acp.2983

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2005). Evaluating Three Criteria for Establishing Cue-Search Hierarchies in Inferential Judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 1088-1104. DOI: 10.1037/0278-7393.31.5.1088

Ramirez, J. C. & Marshall, J. A. R. (2017). Can natural selection encode Bayesian priors? Journal of Theoretical Biology, 426, 57-66. DOI: 10.1016/j.jtbi.2017.05.017

Rachlinski, J.J. & Wistrich, A.J. (2017). Judging the Judiciary by the Numbers: Empirical Research on Judges. Annual Review Of Law And Social Science, 13, 203 – 229. DOI: 10.1146/annurev-lawsocsci-110615-085032

Saks, M.J. & Thompson, W.C. (2003). Assessing evidence: Proving facts, [in:] Handbook of Psychology in Legal Contexts, Eds. Carson D., Bull R., John Wiley and Sons, Chichester, pp. 329 - 345.

Saks, M.J. & Spellman, B.A. (2016). The Psychological Foundations of Evidence Law. New York University Press, New York.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. Law & Society Review, 17, 105-151. DOI: 10.2307/3053534

Schwarz, N. (2007). Attitude construction: Evaluation in context. Social Cognition, 25, 638-656. DOI: 10.1521/soco.2007.25.5.638

Simon, H. A. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69, 99-118. DOI: 10.2307/1884852

Simon. D, (2012). In doubt. The Psychology of the Criminal Justice Process. Harvard University Press, Cambridge, Massachusetts, London, England.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27, 1250–1260. DOI: 10.1037/0278-7393.25.5.1250

Spellman B.A. & Tenney E.R. (2010). Credible testimony in and out of court. Psychonomic Bulletin & Review, 17, 168 – 173. DOI: 10.3758/PBR.17.2.168

Thagard, P. (2005). Testimony, credibility, and explanatory coherence. Erkenntnis, 63, 295 – 316. DOI: 10.1007/s10670-005-4004-2

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute Identification by Relative Judgment. Psychological Review, 112, 4, 881–911. DOI: 10.1037/0033-295X.112.4.881

Trueblood, J. S. & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. Cognitive Science, 35, 1518-1552. DOI: 10.1111/j.1551-6709.2011.01197.x

Trueblood, J. S., Yearsley, J. M., & Pothos, E. M. (2017). A quantum probability framework for human probabilistic inference. Journal of Experimental Psychology: General, 146, 1307-1341. DOI: 10.1037/xgen0000326

Yearsley, J. M. & Pothos, E. M. (2016). Zeno's paradox in decision making. Proceedings of the Royal Society B, 283, 20160291. DOI: 10.1098/rspb.2016.0291.

Yearsley, J. M. & Trueblood, J. S. (2018). A Quantum theory account of order effects and conjunction fallacies in political judgments. Psychonomic Bulletin & Review, 25, 1517–1525. DOI: 10.3758/s13423-017-1371-z

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. Proceedings of the National Academy of Sciences, 111, 9431–9436. DOI: 10.1073/pnas.1407756111

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: the constructive role of articulating impressions. Cognition, 133, 48-64. DOI: 10.1016/j.cognition.2014.05.015

White, L. C., Barque-Duran, A., & Pothos, E. M. (2015). An investigation of a quantum probability model for the constructive effect of affective evaluation. Philosophical Transactions of the Royal Society A, 374, 20150142. DOI: 10.1098/rsta.2015.0142

White, L. C., Pothos E. M., & Jarrett, M. (2020). The cost of asking: how evaluations bias subsequent judgments. Decision, 7, 259-286. DOI: 10.1037/dec0000136

Wilson, W. (1971). Source Credibility and Order Effects. Psychological Reports, 29, 1303–1312. DOI: 10.2466/pr0.1971.29.3f.1303

Wojciechowski, B. W. & Pothos, E, M. (2018). Is there a conjunction fallacy in legal probabilistic making? Frontiers in Psychology, 9, 391. DOI: 10.3389/fpsyg.2018.00391

Appendix A: Legal context

The study was conducted in Poland, with the use of court files from the Polish courts and with participants recruited in Poland. For this reason, we offer a quick overview of the judicial system in Poland. As opposed to the common law system, based on judicial precedents and case law developed over centuries, the Polish law relies heavily on written codes and statues. Precedents hold less weight and the main focus is on interpretation of the laws by legal scholars and judges. Comprehensive legal codes cover various aspects of law, including the criminal law. The criminal process is inquisitorial, where the judge takes a more active role in investigating a case. The judge participates actively in the trial, especially during the examination of the evidence, in accordance with the regulations for criminal proceedings in Poland. The judge may initiate questioning, collect evidence (note that judges can decide to search and collect evidence not provided by either prosecutor or attorney), and direct the proceedings, noting that, from the judge's point of view, the primary goal of the proceedings is to uncover the ground truth, rather than to advocate for one side. The burden of proof lies with the prosecution and guilt has to be proven beyond a reasonable doubt. Defendants are deemed innocent until proven guilty.

Only a public prosecutor has the authority to order a filing of charges, to determine whether pre-trial detention should be used, and how the case should be concluded (i.e., whether to discontinue the proceedings or file an indictment). The accused have the right to a legal advisor, to argumentation, to participate in procedural actions, to appeal against procedural decisions, and to be acquainted with the case files. Accused persons have the right to employ a legal adviser (attorney or solicitor) for their defense; if the accused cannot afford to pay for one, they may be granted legal aid at public expense.

Court procedures typically consist of an oral hearing, that is put on record and is open to the public. Judges and jurors consider and vote on guilt, the legal classification of the crime, and other issues, before the judge decides on sentencing. The process concerning one case is usually concluded in one sitting. If it is necessary to obtain a lot of evidence (e.g., interview witnesses or expert opinions) or the case is complex in some other way, the chairman of the panel organizes the proceedings and sets dates for the hearings. Sometimes the proceedings can go on for several years. The verdict, which addresses both the issue of guilt and the appropriate punishment, is intended to be a cohesive whole. The verdict must first be verbally announced in the courtroom by the presiding judge, before put in writing.

According to the rules of conducts, the judge and the jury members must consider three factors when determining the sentence: 1) the evidence and how it was evaluated, 2) the rules of logic, and 3) their own knowledge and experiences. Although a suspect's confession is seen as an important piece of evidence, it is generally insufficient to establish guilt by itself. There are no distinct evidence presentations made by the prosecution and the defense throughout the trial, unlike in the common law system. The court must evaluate expert testimony in accordance with the rules for evidence set out by the Polish criminal procedures. Therefore, an expert's determination is not binding for the court. However, as for many other judicial systems, courts will accept expert testimony, particularly if provided by psychiatrists. This is especially true if the criminal's mental state or potential for threat are relevant.

Judges review a file for a case prior to the formal trial, which is prepared by the prosecutor. Therefore, the information in the case file would typically be intended to incriminate and a judge's initial impression is probably biased towards guilt. The file might include information indicating innocence too. During the subsequent court proceedings, the prosecutor and defense attorney attempt to influence the judge and it is up to the judge how to weigh the corresponding evidence.

Judges in Poland are impartial and answerable only to the law. The mandatory retirement age is 70 and the minimum age for appointment is 26. Candidates for judicial appointments must be employed as assistant judges for at least two years and pass a public exam.

Appendix B: additional notes on the statistical model

To remind you, the design of the study was 2 (evidence order: guilty-innocent, innocent-guilty) x 2 (rating condition: single vs. double rating, to mean either one rating for both pieces of information at the end or a preliminary rating after the first piece of information followed by a final rating) x 4 (role: judge vs prosecutor vs attorney at law vs layperson) x 6 (case; there were six cases) mixed effects. Note, it is appropriate to model 'case' as a fixed effect, since the six cases were specifically chosen with a view to make the Evaluation Bias and order effects plausible, as is standard practice in research on decision fallacies (e.g., Tversky & Kahneman, 1983; Huang et al., in press). The goal of the present research was to offer an existence proof of the Evaluation Bias and order effects, in a sample comprised of legal professionals, not to claim that such biases occur generally with other case materials (again, by analogy with decision research on fallacies).

A prerequisite concerning the assessment of evidence for the four fixed effects (evidence order, rating condition, role, and case) is to identify a suitable linear model for the data. We proceed based on the practice of starting with a basic model (the simplest possible model) and gradually elaborating this (through the addition of random coefficients; Field, 2017; Raudenbush & Bryk, 2002; Twisk, 2006; but see Barr et al., 2013; Winter, 2013, for arguments that all random slopes justified by the experimental design should be included). Evidence for more elaborate versions was assessed using a Maximum Likelihood Estimation (MLE) method. Model fits were expressed by minus twice log likelihoods (-2LL) and nested models were compared using chi squared distributions for the difference in model fits, with degrees of freedom corresponding to the difference in model parameters. The significance level is taken to be .05.

We examined a series of nested models to establish, first, which terms should be retained, second, whether random effects had to be modeled with just intercepts or with intercepts and slopes and, third, if slopes had to be included, the structure of the covariance matrix (Field, 2012). The table below (Table B1) shows the results of this exercise. Using Wilkinson notation (Wilkinson & Rogers, 1973), the supported model was order*rating*role*case+1|participants. Adding random effects for slopes (with a simple variance components covariance matrix) worsened fit compared to the previous model, so we stopped elaborating the model at that point. The table below shows information for the models we examined.

model	parameters	"-2LL"	chisquare p-value for present model, against best previous model	conclusion
order+rating+1|participants	3	308		
order*rating+1|participants	6	269	<0.05	Evidence for interaction between order, rating
order*rating*role+1|participants	18	256	0.37	No evidence for role by itself
order*rating*role*case+1|participants	97	-85	0.00	Evidence for role and case, together with main fixed effects
order*rating*role*case+(1+order+rating+role+case|participants), variance components	101	-77	NA	No evidence for adding slopes

Table B1. Identifying the best statistical model for ratings of guilt.

Appendix C: Attempt to recast the study with a within participants design

To summarize from main text, there were two purposes to this pilot. The first purpose was to replicate the order effect and Evaluation Bias findings, from the main experiment, but in controlled experimental conditions. If a replication is observed, then this would mitigate concerns that participants did not follow the procedure in the main experiment as intended. The second purpose was to expand the test of the association between order effects and the Evaluation Bias, by having measures for both effects within participants. A limitation of this follow-up study was that it was not possible to recruit legal professionals, rather we had to rely on university students.

Recall, in the main experiment each participant received the six legal cases, with a random combination of factor levels for each case (IG or GI order; single or double rating condition). To accomplish a fully within participants design, we would need four case iterations for each participant. We considered this implausible, given the memorability of each case. So, we settled with having two presentations of each case for each participant, with a change in either order or rating condition in the second presentation. The two presentations were in separate testing sessions, separated by three weeks. The design is still partial, but less partial compared to the main study.

To anticipate our conclusions, the first purpose of this follow-up was accomplished (restricting the data to the first presentation, both order effect and Evaluation Bias were replicated), but not the second (including results from the second presentation, the Evaluation Bias disappeared, but evidence for order effects was still apparent).

Design and participants

The design is identical to that of the main study, but without the fixed effect of role (all participants were university undergraduates). There were three fixed effects of rating condition (single vs. double), order (GI vs. IG), and case (six levels, corresponding to the six cases) and a single random effect of participants. We analyzed separately results from the first and second presentation stage.

Study participants were 52 psychology students (39 women, 13 men) from the Institute of Applied Studies at the Jagiellonian University in Krakow. The participants were taking one of the courses taught by the first author. All 501 students assigned to first author's teaching groups were invited to take part in the study. Participants' ages varied between 19 and 27 years (M = 22.46 years; SD = 1.71). Participating in the study was voluntary. Subjects were informed that the aim of the research was to study application of the quantum probability theory to decision-making. Participants received no course credit for taking part in the research, but those who participated in both stages of the study received a $10 voucher for a books and stationery store.

Ethics approval for this work was provided by the Ethics Committee of the Faculty of Management and Social Communication at the Jagiellonian University and all participants provided written consent prior to participation.

Materials and procedure

The materials were identical to those of the main study. Regarding procedure, in each of two stages participants were given a booklet with the six legal cases (as in the main study) and were asked to read the instructions, then evaluate the legal cases. Participants were tested in small groups (3 – 8), so that it was straightforward to assess whether they processed the legal cases in the intended way. Testing took place in one of the classrooms at the Institute of Applied Psychology, at the Jagiellonian University. Participants could arrive any time at the classroom, between 10:00 and 13:00, on particular days. Once seated and given the materials, they were told they could go through the cases at their own pace. The study typically lasted between 15 to 20 minutes.

In the second presentation stage, after three weeks, each of the returning participants would receive the same six cases (the names were altered as an attempt to reduce superficial similarity, but many participants reported remembering the cases). For each case, we would randomly switch one factor (the rating or the order factor), compared to what it was for that participant in the first presentation.

Results

We employed the same statistical model as for the results of the main study, separately for the first and second presentation stage. Note, as before, random effects were modeled only with intercepts.

Regarding the first presentation stage, the model had a -2LL of -96. The order effect was significant ($F(1,288)=10.3$, $p=.001$), as well as the interaction between order and rating condition, which is the Evaluation Bias ($F(1,288)=8.3$, $p=.004$). All interactions with the case fixed effect were significant (worst p-value .003). As Figure 1C shows, the broad trends are as in the main experiment.

Regarding the second presentation stage, the model had a -2LL of -104. The order effect was significant ($F(1,228)=4.3$, $p=.04$), but not the interaction between order and rating condition – that is, there was no evidence for the Evaluation Bias in the second stage ($F(1,288)=.004$, $p=.9$). All interactions with case were significant (worst p-value .04). Looking at the first vs. second presentation stage, there was a trend for judgments of guilt to be lower (.52 vs .48, $F(1,557)=3.3$, $p=.07$). Several participants did report remembering the legal cases between the two presentation stages, however, we did not collect recognition data.

Finally, we checked the case-based preliminary examination of the association between the Evaluation Bias measure and the order effect (both calculated as in Section 2.5.3). For the data in the first phase, the second stage, and together, the correlations were respectively 0.46 (NS), 0.96 (p=<.0001), and 0.96 (p=.003). Again, we caution about the robustness of these correlations given the small number of items.

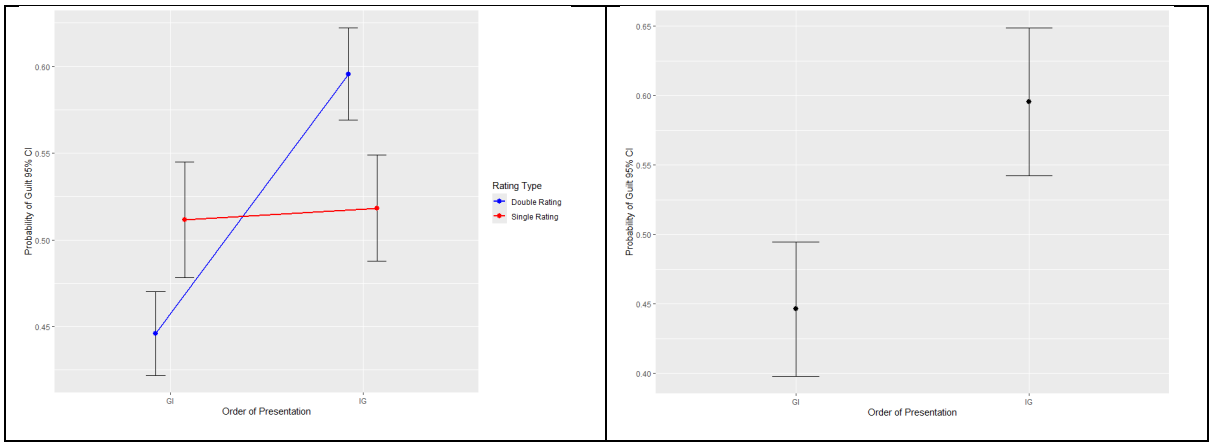

Figure 1C. Overall mean participant final ratings of single and double rated GI and IG stimuli (left panel) and order effects (right panel).

Discussion

This attempt at a partial within participants design showed that, in the first presentation stage, both the order effect and the Evaluation Bias replicated, as in the main study. However, in the second presentation stage, the Evaluation Bias did not replicate. The entire premise of the Evaluation Bias is that there is a difference between just observing some information vs. observing the information and making a corresponding judgment. With multiple presentations of the same legal case, this logic is undermined. We hoped that separating the two presentation stages by several days might partly mitigate the problem, but this was not the case. It is unclear to us how we can accomplish a within participants design, with materials like the present ones (that is, materials which are highly distinctive and fairly memorable, and for which it is not possible to procedurally generate arbitrary variations). This remains a challenge for future work.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Field, A. P. (2017). Discovering Statistics Using IBM SPSS Statistics (5th edition). Sage.

Field, A. P., Miles, J., & Field, Z. (2012). Discovering statistics using R. Sage.

Huang, J., Busemeyer, J. R., Ebelt, Z., & Pothos, E. M. (in press). Bridging the gap between subjective probability and probability judgments: the Quantum Sequential Sampler. Psychological Review.

Raudenbush, S. W. & Bryk, A. S. (2002). Hierarchical models (2nd edition). Thousand Oaks, CA: Sage.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunctive fallacy in probability judgment. Psychological Review, 90, 293-315.

Twisk, J. W. R. (2006). Applied multilevel analysis: a practical guide. Cambridge: Cambridge University Press.

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic Description of Factorial Models for Analysis of Variance. Journal of the Royal Statistical Society. Series C (Applied Statistics), 22(3), 392–399. https://doi.org/10.2307/2346786

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. https://doi.org/10.48550/ARXIV.1308.5499

Supplementary Electronic Material 1

We present translations of the summaries of the legal cases used in the experimental work. For each case, we also show absolute mean difference between IG and GI orders and standard deviation (note, the absolute value was computed, after averaging across participants). The size of this difference is an indication of how 'non-classical' each of these cases is.

We briefly comment on the cases for which there was the highest (MS) and lowest (AP) order effect respectively. In the case with the highest order effect (biggest difference), in the one part of the description it is established that the suspect possessed an illegal substance and the suspect confesses and claims that he had possessed amphetamine. In the second part of the description, it is stated that according to the results of laboratory test the substance was not amphetamine, but a mixture of salt and caffeine. So, one part indicates guilt and the other part innocence. We can plausibly imagine that, given the conflicting nature of the two parts, even a slight recency bias would be amplified to a large order effect.

In the AP case, both parts indicate that the crime was committed but the information included in one part indicated that there were two perpetrators, but the information in the other part that there was only one. So, the difference between the two parts was not about whether the crime was committed or not, but rather whether the suspect AP had committed the crime by himself or with an accomplice. The two parts are not really in opposition, suggesting that it is easier to consider the information together (cf. Pothos et al., 2017). Note also that raters reported finding this case a bit difficult, because the names of the suspects and victims were similar (Artur P., Andrzej P., Adrian P.). Perhaps the longer time spent processing the two parts contributed to having a more classical representation for the relevant information.

Proceedings against Lucjan L. Mean order effect (IG-GI) = .21, SE = .04, N_{IG}=56, N_{GI}=64

Introduction. Lucjan L. is suspected of submitting, on June 24, 2003 in A., as a genuine fake by an unknown person, a statement of professional preparation to perform independent technical functions in construction, with the registration number 1226/00/AB allegedly issued by the Provincial Office in A. on February 24, 1999, that is, an offense under Article 270 § 1 of the Criminal Code.

Part I. During the proceedings, the suspect Lucjan L. did not admit to committing the alleged act. Initially, he exercised his right to refuse to give an explanation. During a subsequent interrogation, he did not admit to the alleged act and explained that in 1984 he received a decision from the provincial office regarding his professional preparation to perform independent technical functions

in construction. At that time he submitted the required documents - a certificate of completion of a construction technical school and an application for an apprenticeship. Based on the documents submitted, he received a decision on his ability to perform independent technical functions in construction.

The suspect explained that after using the document for about 10 years, the decision was destroyed, so he presented the destroyed decision and applied for a new decision. He received the new decision about 10 years ago, and in 2003 he applied for registration as a member of the Chamber of Civil Engineers, paid the required fee, and submitted a statement of professional preparation previously issued by the provincial office.

Documentary evidence gathered, particularly that sent by the council of the A. District Chamber of Civil Engineers, confirms that the suspect applied for inclusion in the list of engineers and paid the fee required by the regulations.

Part II. In the course of the investigation, a forensic document expert opinion was obtained. The expert opinion shows that the signature under the application for registration as a member of the Chamber of Civil Engineers belongs to the suspect Lucjan L.

At the same time, it was found that the suspect is not listed in the register of persons holding construction licenses. After verification of the documents, it was established that the entries in the 1984 register were completed with the number 744, while the decision submitted by Lucjan L. was numbered 1226. In addition, there was an additional designation "AB" on the document, while no such designations were used in the register numbers of documents issued in 1984.

However, the testimony of witnesses - employees of the provincial office - and the graphological opinion show that the signature on the document stating professional preparation for independent technical functions in construction does not belong to the official responsible for issuing these certificates. The handwritten notation "Zygmunt K." was not written on the document by the head of the Architecture and Landscape Department of the Provincial Office of the city of A., nor by any of the persons employed at the Office at the time.

It was further found that the header seal and the round seal affixed to the document did not conform to the model seal used at the time the statement was allegedly issued at the provincial office.

Proceedings against Mateusz S. Mean order effect (IG-GI) = .29, SE = .05, N_{IG}=59, N_{GI}=61

Note this is the highest order effect.

Introduction. Mateusz S. is suspected of possessing a psychotropic substance in the form of amphetamine in the amount of 0.43 grams on June 13, 2014 in the city of T., at the A. Street,

contrary to the provisions of the law, with the act constituting an incident of lesser gravity, i.e. an offense under Article 62 (1) and (3) of the Act on Counteracting Drug Addiction of July 29, 2005.
Part I. On June 13, 2014, a search was conducted on Mateusz S. A loose substance of white and yellow color placed in a paper bundle was revealed and secured. It was checked using a NARKO 2 drug tester with MARQUIS reagent, which showed that it was most likely a psychotropic substance - amphetamine.

In connection with the above, Matthew S. was arrested, and then a decision was issued against him to present a charge.

During the interrogation, the suspect admitted to the alleged act and gave explanations. He stated that on June 13, 2014, at around 5:00 p.m., from a man he met by chance with the nickname "Hary", he purchased amphetamine for the amount of PLN 30. He stated that he only knows the man by sight and knows that they call him "Hary," but he does not know where he lives and has had no other contact with him. He explained that "Hary" asked him about whether he would like to buy amphetamine from him. He agreed because he wanted to try what it was like after taking it. He reported that he had never bought any drugs before, it was a one-time situation.
Part II. In the course of the investigation, an expert employed by the Physicochemical Laboratory of the Forensic Laboratory of the Regional Police Station in X. was appointed and the secured substance was tested.

The tests conducted did not confirm that the secured substance was a psychotropic substance of the amphetamine group. Instead, they showed that it was α - PVP salt and caffeine, which are not on the list of narcotic drugs, psychotropic substances, and precursors, under the Act on Counteracting Drug Addiction of July 29, 2005.

Proceedings against Jarosław O. Mean order effect (IG-GI) = .03, SE = .04, N_{IG}=69, N_{GI}=51

Introduction. Jaroslaw O. is suspected of the fact that on October 30, 2014 in A. acting jointly and in concert with another identified person, in order to force Irena P. to repay a debt, i.e. a loan in the amount of no less than PLN 1,420, he threatened to use violence, i.e. beatings and bodily harm. The threat aroused a reasonable fear that it would be fulfilled, i.e. an act under Article 191 § 2 of the Criminal Code.
Part I. The suspect did not admit to committing the alleged act. Jaroslaw O. explained that he has been in the business of providing loans for ten years. He granted loans in amounts ranging from PLN 500 to PLN 1,000. He concluded contracts for one month and charged a 15% commission on the loan amount granted. When the borrower defaulted during the settlement period, he concluded a new agreement, which was for the amount of the outstanding loan and interest if not paid before.

Jaroslaw O. did not deny that a brawl occurred between him and Irena P., but claimed that he did not threaten Irena P. He explained that he asked to be accompanied by a colleague because he feared for his safety in a situation where he was dealing with borrowers who did not want to settle their obligations, and the presence of a colleague was not intended to put pressure on the victims.

In the course of the proceedings, no witnesses were identified to corroborate the version of events presented by Irena P. and her partner Bartholomew C. At the same time, it is undisputed that these individuals did not fulfill the contracts concluded with Jaroslaw O.

In the course of the proceedings, several hundred loan agreements were produced that Jaroslaw O., who kept detailed records in this regard, b granted loans to dozens of people. To date, there has been no record from his borrowers (with the exception of Irena P.) of any other criminal cases.

In addition, two witnesses to the October 30 incident - Catherine W. and Dariusz M. - do not confirm the versions of events described by Irena P. and Bartholomew C. Dariusz M. did not hear the alleged threats at all, and Catherine W. testified that she only heard that Jaroslaw O. was said to have threatened Bartlomiej C., not Irena P., while the victim Bartlomiej C. did not confirm this in his testimony.

Part II. Irena P. testified that in 2011 she borrowed from Jaroslaw O. the amount of PLN 600, which she has been repaying until now, with an amount of PLN 1,200 still outstanding due to accrued interest. She testified that until September 2014, she regularly repaid the debt. In October, Jaroslaw O. was said to have firmly demanded the repayment of the money, until a situation arose in which he arrived with two men. In the presence of her partner, Jaroslaw O. was said to have threatened Irena P. with violence if she did not return the money and that he would "break her bones."

A witness to the incident was the victim's partner Bartlomiej C., who confirmed Irena P.'s version of the event. Moreover, as Irena P. and Bartlomiej C. testified in unison, Jaroslaw O. allegedly behaved aggressively using vulgar slurs against Irena P

Proceedings against Artur P. Mean order effect (IG-GI) = .00, SE = .05, N_{IG}=57, N_{GI}=61

Note this is the lowest order effect.

Introduction. Artur P. is suspected of having, on May 2, 2010 at 7:30 p.m. in B. in the area of a pond near C. Street, jointly and in agreement with Marcin G., taking advantage of Adrian P.'s inattention, taken an ABC cell phone with IMEI number 111111 and a SIM card for the purpose of appropriation. He caused losses in the amount of PLN 300 to the detriment of Andrzej P., i.e. a crime under Article 278 § 1 of the Criminal Code.

Part I. Witness Adrian P. testified that on May 2, 2010, he and his friend were at a pond near C. Street where they were fishing. Adrian P. put down his ABC cell phone, which belonged to his father, Andrew P., on the grass a short distance away. At some point, he noticed Marcin G. and Artur P. walking nearby. When he turned around, he noticed that Marcin G. was holding his cell phone. Adrian P. started to chase Marcin G., but the perpetrator managed to escape.

Marcin G. faced charges of committing a crime under Article 278 § 1 of the Penal Code. Questioned as a suspect, he did not admit to committing the alleged act and indicated that the theft was committed by Artur P.

Part II. Artur P. was charged with committing a crime under Article 278 § 1 of the Criminal Code in cooperation with Marcin G. Questioned as a suspect, he did not admit to committing the alleged act and explained that he had nothing to do with the theft of the phone.

On July 22, 2010, the victim Andrzej P. notified the Z. District Police Station that the stolen phone had been returned to him by Marcin G.

Witness Adrian P., son of the victim Andrzej P. testified emphatically that the phone was stolen by one person.

Proceedings against Jakub P. Mean order effect (IG-GI) = .11, SE = .05, N_{IG}=56, N_{GI}=63

Introduction. Jakub P. is suspected of providing a minor, Tomasz K., with a narcotic drug in the form of cannabis herb other than fibrous hemp in the amount of 0.15 grams for the amount of PLN 15, in the period from 1 to 7 July 2012 in L., in order to gain financial gain, i.e. an offence under Article 59(3) in connection with Article 59(2) of the Act of 20 July 2005 on counteracting drug addiction.

Part I. On 14 July 2012, officers of the District Police Headquarters in B., in the course of performing their duties, noticed a young man who, upon seeing the patrol, started to behave nervously and quickly put an object in his pocket. As a result, the officers proceeded to identify the man, who turned out to be Tomasz K. During the search, a string bag with green-coloured dried plants was revealed.

In order to determine the type of the secured substance, it was tested with the Narko 2 drug tester. As a result of the test, the tester turned red, indicating that the seized substance was cannabis.

In order to confirm the test results, an expert from the Forensic Research Centre was consulted by order of 6 September 2012, who, in an opinion issued on 19 September 2012, stated that the green-brown coloured dried plant with a net weight of 0.15 g is cannabis herb other than fibre.

When questioned as a witness, Tomasz K. testified that he bought the narcotic drug in the form of cannabis for PLN 15 from Jakub P. It has been established that at this time Jakub P. was on holiday in the city of B.

Part II. When questioned as a suspect, Jakub P. categorically denied ever having sold cannabis to Tomasz K. He could not explain why Tomasz K. had named him as the dealer.

Questioned again as a witness, Tomasz K. (in the course of an indirect confrontation) testified that it was indeed not true that Jakub P. had sold drugs to him. He added that he named Jakub P. because he was scared that he would be punished. He added that he found the cannabis near a bench in the school playground.

Proceedings against Krzysztof B. Mean order effect (IG-GI) = .19, SE = .05, N_{IG}=51, N_{GI}=67

Introduction. Krzysztof B. is suspected of driving, on 5 December 2012 in B. on a public road in B. Street, a Z. car with registration number ABCXXX in a state of intoxication. The control and measuring device showed 1.39 mg/l of ethyl alcohol in the breath, i.e. an offence under Article 178 a § 1 of the Penal Code.

Part I. On 5 December 2012, officers of the District Police Headquarters in B.: Szymon J. and Jakub M., while on their road patrol, noticed in B. Street in B. a Z. car with registration number ABCXXX, which was in a roadside ditch. The officers, with the door of this vehicle open, also noticed a man exiting the car, who turned out to be the owner of the vehicle, Krzysztof B.

Officers assisted Krzysztof B. out of the ditch onto the road. Due to the fact that the owner of the vehicle said that he was driving the vehicle and due to the fact that he could smell alcohol, he was subjected to breath alcohol tests. Szymon J. and Jakub M. called a police car from the traffic section to the scene.

Tests conducted on the spot showed that Krzysztof B. was in a state of intoxication at the time of the test, as the control and measuring device showed 1.39 mg/l of ethyl alcohol in the air he was breathing out.

Part II. Krzysztof B. was charged. The suspect did not admit to committing the alleged act and explained that on 5 December 2012 he was at the construction site of his house in B. where he arrived in his Z. car with registration number ABCXXXX. At the construction site, he consumed about 5 beers. He then asked his friend Mariusz M. to drive him home in his car. Mariusz M. agreed, but after leaving the site after approximately 500 metres, he skidded and the vehicle rolled into a ditch. Mariusz M. got out of the car from the driver's side and went to retrieve his vehicle and was told to call for roadside assistance. By this time, Krzysztof B., sitting in the front passenger seat, had managed to exit the vehicle, and police officers had already arrived at the scene.

Questioned as a witness, Mariusz M. confirmed the suspect's explanations and testified that he had arrived on 5 December 2012 at a construction site in B. where he was supposed to help with construction work. At the site, he found Krzysztof B., who was intoxicated and asked to be driven home in his Z. Mariusz M. was reluctant to drive the suspect in his car because the vehicle had an automatic transmission and he had never driven such a vehicle. Due to Krzysztof B.'s insistence, Mariusz M. gave in. He got behind the wheel, took Christopher B. as a passenger who sat in the front and then started driving the vehicle, driving out of the property. About 600 metres further on he got into a skid. As he testified, it was slippery and, in an attempt to brake, he pressed on the accelerator instead of the brake, as a result of which he lost control of the car and it rolled into a ditch. An argument ensued between the two due to Christopher B.'s claims against him for mishandling the vehicle. Mariusz M. angrily returned to the construction site where he had parked his car and approached Krzysztof B. to help him, but the latter was upset and continued to insult him. Mariusz M. then drove away from the scene.

The witness Zbigniew M., who was heard twice in the case and who carried out construction work at the site, testified consistently that he saw Krzysztof B. take the front seat in the passenger seat and another man sit behind the wheel; and then both men got into the Z's. car, which was parked on the premises in B. Street. They then drove off together in vehicle a Z. belonging to the suspect. Witness Zbigniew M. observed this after coming down from the scaffolding when he was on the ground.