Adjusting Legal Standards

Shay Lavie, Tal Ganor, Yuval Feldman[*]

Abstract: This paper seeks to explore whether the interpretation of legal standards is influenced by decision-makers' substantive decision. Prior literature on motivated reasoning has shown that decision-makers "shift" their perception of evidence in their desired direction. To the extent this logic applies to legal-standards, we should expect decision-makers to adjust the perception of the legal standard accordingly—e.g., one's decision to favor the plaintiff would induce a pro-plaintiff interpretation of the required threshold to win a case. We present the results of two experiments in which we asked subjects to report their interpretation of the applicable legal threshold after deciding a case, under different legal thresholds. Our participants, by and large, did not shift the legal standard to conform to their substantive decision, contrary to the theoretical expectations. We thus conclude that decision-makers treat the legal standard distinctly than regular evidence.

Keywords: Empirical legal studies, Legal thresholds, Coherence-shifts, Dismissal standards, *Twombly*

JEL Codes: *K40*. *K41*. *D91*. *C91*.

[*] Assistant Professor, Tel Aviv University, Law School; Ph.D. Candidate, Hebrew University of Jerusalem, Psychology Department; Mori Lazarof Professor of Legal Research, Bar Ilan University School of Law. This is a post-peer-review, pre-copyedit version of an article that will be published in the European Journal of Law and Economics. The final authenticated version is available online at the following link.

We thank Ilana Ritov, Dan Simon, Theresa Squatrito, Eyal Zamir, and participants in the tribute workshop in honor of Theodore Eisenberg, the Italian Society of Law and Economics annual conference, and the Conference on Empirical Legal Studies in Europe for useful discussions and comments. We thank Ofir Givaty, Daniel Fishelovich, and Sharon Salinas for excellent research assistance. This research project was supported by the Israeli Science Foundation (grant No. 1251/15).

1. Introduction

a. Motivation and theoretical background

"More probable than not." "Plausible on its face." "Clear and convincing evidence." "Beyond reasonable doubt." "Nonfrivolous argument." These are vague maxims, but fact-finders are commonly asked to interpret these and other legal standards in order to decide a case. This paper seeks to explore whether, and to what extent, the interpretation of legal standards is influenced by decision-makers' substantive decision. For instance, one's decision to favor the plaintiff could bias her interpretation of the required threshold to win a case. The interdependence between the merits decision and the interpretation of the legal rule is an important question. Decision-makers are often asked to first decide and then state their interpretation of the applicable legal standard, and subsequent decision-makers rely on previous interpretations of those standards. A biased interpretation of legal thresholds could thus sway future decision-making. The importance of this question notwithstanding, we found in the literature scarce discussion thereof.

One can think of two general theoretical perspectives to analyze the interdependence between merits decisions and interpretation of legal thresholds. First, the legal threshold could be conceived as any other piece of evidence. In that case, the interpretation of the legal rule is plausibly susceptible to the influential coherence-based model of reasoning (Simon 2004; Engel 2009). Under this model, decision-makers "adjust" the evidence to cohere with their conclusion—evidence are treated "like the pieces of a jigsaw puzzle," and decision-makers "try to fit as many pieces as possible" (Engel 2009, p. 454). Supporting evidence is strongly endorsed while unfavorable evidence is dismissed or ignored. This process turns difficult decisions to coherent ones, as the "unconscious transformation" of the evidence leads "to a seemingly straightforward choice between a compelling alternative and a weak one" (Simon 2004, p. 513). The shift in the perception of the evidence thus results in a more polarized decision-making and a greater confidence in one's conclusion.[1]

[1] The basic paradigm which explains how people motivational needs, self-interest or coherent, affect decision-making is Motivated Reasoning (Kunda 1990).

It is natural to assume that a similar logic would apply to legal standards (Scurich 2012, pp. 64-65, 81). After all, the legal standard, together with the evidence, generate the outcome. To the extent legal standards are treated similarly to other evidence, then, we should expect the perception of legal standards to shift as well. Decision-makers who decide for (against) the plaintiff will report a lower (higher) threshold, making their general conclusion easier. Indeed, a similar phenomenon seems to occur in the real-world, where judges "pad" their decisions with legal precedents that support the desired conclusion (compare Simon 2004, p. 549). This is, then, a "coherence-based" view of legal standards.

On the other hand, legal standards are commonly thought of as materially different from other pieces of evidence—what we can term as a "legalistic" view of standards. This view emphasizes the bifurcated nature of legal decision-making. According to this common description, the decision-maker processes the evidence first; then, she compares the facts to the standard (Clermont 2009). The unconscious, coherence-based reasoning could still take place at the first step, where the decision-maker evaluates the evidence and forms its assessment of the case. However, to the extent the standard is exogenous to the consideration of the evidence, decision-makers perceive the legal standard only after the (coherence-based) formation of one's mind regarding the evidence. This logic suggests that it would be harder for decision-makers to implement the unconscious coherence-based model with regard to the operation of the legal standard. Thus, under the two-step, legalistic view decision-makers are less likely to "adjust" the standard in the direction of their substantive judgment. As a corollary, the description of the legal standard as an independent operator also suggests less polarization, e.g., between those who decided for and against the plaintiff, and weaker confidence in one's overall decision.

The foregoing two, conflicting predictions, then, are based on the likely manner in which decision-makers effectuate the standard of proof. Do they treat it as other pieces of evidence? Or, does the standard embody a distinct cognitive judgment, which follows the assessment of the factual background? Will decision-makers state a different threshold, which conforms to their merits decision? What is the extent, if any, of this "shift"?

In an attempt to start answering these questions, we present the results of two experiments in which we asked subjects to report their interpretation of the applicable legal

threshold after deciding a case (in numerical terms, on a 0-100 scale). Our participants decided a legal dispute in two points in time, t1 and t2. At t1, participants had rudimentary information of the case and were asked to decide whether to dismiss it at the outset or not, under a relatively low legal threshold. This decision-making step was designed to mimic the current trend in American law to enable judges to dismiss, soon after filing, cases that appear to be inappropriate (*Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 [2007]). At t2, our participants received more information, and they were asked to decide whether to fully accept the case under the regular threshold to win a case in common law, the preponderance of the evidence rule.

b. Main findings

Our main findings are the following. First, and most importantly, by and large participants did not shift the legal standard to conform to their substantive decision. More precisely, we did observe some coherence-based reasoning of legal standards at t1 (the dismissal stage); but not at t2 (the final decision). By contrast, at t2 our participants did shift their perception of the *evidence*. The following graph demonstrates the perception of the evidence (the two black slopes) and the interpretation of the legal standard (the two green slopes), both on a 0-100 scale (the y-axis). The x-axis shows the change from t1 to t2, where those who decided for the plaintiff (defendant) are represented by a solid (dashed) line:[2]

[2] We discuss these findings in detail below. *See infra* subsection 4.a. and Table 5.

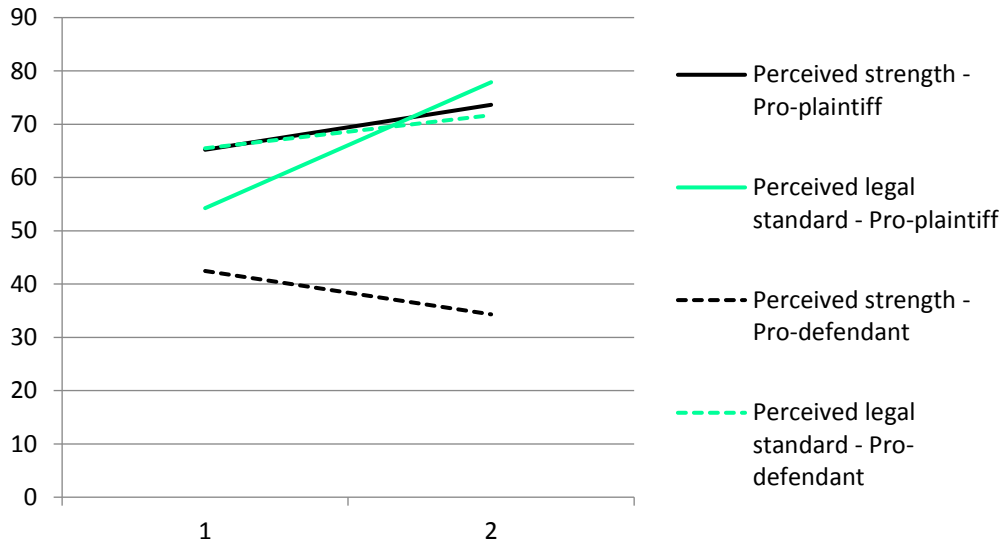

Figure 1. Perceived strength and legal standard on a 0-100 scale (y-axis), at t1 and t2 (x-axis).

One can observe that, with more information at t2, the perception of the evidence has become polarized, as evidenced by the two black slopes. The pro-plaintiff (solid black line) rated the case's strength higher at t2 than at t1; and the pro-defendant (dashed black line) rated the case's strength weaker at t2 than at t1. This is, of course, consistent with the coherence-based model of reasoning (Simon 2004). One would expect the perception of the legal standard to follow a similar pattern. The perception of the standard (two green slopes) of course rose from t1 to t2, as the legal standard at t2 is more demanding. However, the pro-plaintiff (solid green line) and pro-defendant's (dashed line) perceptions have not diverged throughout the move from t1 to t2—rather, they have actually become closer to each other. Participants, then, have not moved their perception of the standard in the direction of their final decision, suggesting that legal standards are treated materially different from other evidence, consistent with the two-step, legalistic description.

An additional finding is a counter-intuitive tendency by some of our subjects to shift the standard in a reverse direction at t2. More precisely, those who decided to reject the case reported a *lower* standard for winning than those who decided to accept. This move is inconsistent with the coherence-based perception of legal standards. Why should decision-makers report a standard that does not cohere, and even conflicts, with their conclusion? We hypothesize that such a "reverse-shift" is a potential tool to justify one's conclusion—a lower standard for winning could indicate to the pro-defendant decision-maker the

strength of its evidence-based position, a logic somewhat akin to self-signaling (Bénabou & Tirole 2006). Indeed, these findings mainly stem from a group of subjects who decided a conflicting decision (for the plaintiff) at t1, and hence were under the greatest pressure to justify their decision at t2. This phenomenon is consistent with a "legalistic" view of legal standards—decision-making is a two-step process, where the legal standard could justify a substantive decision.

Our findings regarding the nuanced interpretation of legal standards, depending on one's substantive decision, have various legal implications. In general, the interdependence between legal standards and merits decisions advises caution regarding our reading of interpreted legal thresholds. We end our paper with a possible actual example. Following concerns regarding frivolous litigation and costly proceedings, judges in the U.S. are now guided to dismiss, at the preliminary stages, cases that do not pass a low threshold (*Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 [2007]). To the extent judges shift the interpretation of this dismissal standard, and subsequent judges rely on past interpretations, dismissal standards should be biased downward over time, rendering them ineffective.

c. Prior literature and organization of the paper

The general goal of this paper is to marry two bodies of experimental literature. The first, which we elaborated on above, concerns the ways in which decision-makers "adjust" their perception of the evidence, in order to generate a more coherent judgment (Simon 2004; Engel 2009). The second explores the interpretation and implementation of legal thresholds (Kagehiro 1971; Kagehiro & Stanton 1985; Zamir & Ritov 2012; Woody & Greene 2012). We found in the literature scarce discussion on the interdependence between legal threshold and merits decisions, and little experimental evidence of shifts in the perception of legal standards (Phillips 2002; Glöckner and Engel 2013). As we discuss in greater detail in Sections 3 and 4 below, we contribute to the literature by presenting evidence of some coherence-shifts of legal standards at t1, but not at t2; by introducing the distinction between preliminary (some shift in the standard) and final stages (no shift in the standard) of the legal process; and by pointing to a possible reverse-shift in the perception of the legal standard at t2 among those who made a conflicting decision at t1.

Beyond the interdependence between merits decisions and interpretation of legal thresholds, our paper contributes to the body of literature that looks at the ways in which legal standards are translated into numerical terms. A salient finding is that legal standards are interpreted as higher than they "should." To illustrate, the well-known common-law threshold to win in civil cases, the preponderance of the evidence rule ("more probable than not"), is assumed to equal 50% in theory, but in actuality is translated considerably higher, to around 75% (Simon & Mahan 1971; Glöckner and Engel 2013, p. 242). Our findings are consistent with this literature, as we find that participants translate the preponderance rule to 75%-80%. We contribute to this branch of literature by exploring the interpretation of legal thresholds to dismiss a case at the outset. We find that the numerical interpretations of dismissal standards are similarly exaggerated, consistent, of course, with the general trend to inflate numerical thresholds.

More broadly, our paper is also related to bodies of literature that investigate legal standards and, more specifically, burden of proof. Rich literature has discussed the use of open-ended legal standards, such as "reasonable care," rather than bright-line rules. By and large, the former offers wider discretion in its implementation to the specific circumstances; the latter requires more effort to enact, but better guides primary behavior (e.g., Kaplow 1992; Johnston 1991; Fon and Parisi 2007). While this literature focuses on legal standards that guide primary conduct, our interest is in the interpretation of legal standards that are designed to specify the evidential burden of proof to proceed, e.g., "more probable than not." Prior literature has shown that the level at which the probabilistic threshold of proof is set has subtle effects on primary conduct, i.e., deterring wrongdoing but chilling beneficial behavior (Kaplow 2012); and litigation behavior (Guerra et al. 2018)

Finally, the findings in our paper stem from experiments in which participants were asked to decide whether to dismiss the case at the outset and/or rule for the plaintiff at the end of the proceedings. These two decisions are made at different times, with varying evidentiary backgrounds, and under different legal standards. The gradual decision-making process in our experiments was designed to mimic the relatively new dismissal standards in American law (*Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 [2007]). A considerable body of literature discusses the theoretical and empirical repercussions of the new, more searching dismissal standards (Bone 2009; Rachlinski 2011; Engstrom 2013; Hubbard

2016). Despite the growing importance and interest in dismissal standards, to the best of our knowledge this is the first study that tests the *Twombly* dismissal standards in an experimental setting. Our results can thus shed light on the actual implementation of the *Twombly* standards in American law, a point that we briefly discuss in our concluding section.

The paper proceeds as follows. Section 2 presents our methodology and the baseline results regarding the numerical interpretation of legal standards. Section 3 presents the numerical interpretation of legal standards among subjects who were asked to decide whether to dismiss at the outset (t1). Section 4 presents the results regarding the final decision (t2). Section 5 discusses the broader implications and concludes.

2. Baseline Numerical Perceptions of Legal Thresholds

We investigated the perceptions of several legal standards in civil adjudication. Specifically, we were interested in the preponderance standard for winning a case in common law ("more probable than not") and the standards that are required to dismiss a case at the outset. In our main experiments we presented participants with a description of a legal case and asked them to decide the case based on a given legal threshold. Following their decision, we asked the subjects to report their perception of the applicable legal standard, in numerical, probabilistic terms, on a 0-100% scale. While it is not trivial to think that decision-makers reason in probabilistic terms (Vars 2010, pp. 18–24), it is common for courts and scholars to presume that legal standards can be expressed as probabilistic thresholds (Engel 2009, p. 440; Pardo 2015, pp. 1827-27).

To capture the interdependence between the numerical interpretation of legal standards and the decision on the merits, we compared these numerical interpretations of legal standards from our main experiments to benchmark numerical perceptions of the same legal standards. We derived the benchmark figures by asking other participants to translate the same legal thresholds to numbers in the abstract, without actually applying these standards to a dispute. We first present and discuss our results regarding the abstract, "naked" perceptions of various civil standards.

To acquire the baseline interpretations of legal standards, we asked participants to evaluate, in probabilistic terms, three different thresholds. The first (and most onerous) is the familiar preponderance of the evidence rule.[3] The second is the *Twombly*-like standard to survive dismissal in American courts—federal courts in the U.S. are guided to dismiss where the plaintiff's claim is not "plausible on its face."[4] The third is a lower standard to survive dismissal—unless the claim seems frivolous.[5] This lower, non-frivolous standard is inspired by the threshold requirements to impose sanctions on American parties (Federal Rules of Civil Procedure, Rule 11(b)).

We surveyed 183 participants regarding these three standards, and we used a 3!=6 group design to eliminate order effects in the presentation of the standards (we found no order effects). We recruited our participants, representative of the general population, through an Israeli online platform.[6] The following table presents our results:

Decision standard	Probabilistic translation
Preponderance	81% (1.6)
Plausibility	68% (1.8)
Non-frivolousness	68% (1.9)

Table 1. Means and standard deviations of benchmark probabilistic perceptions of legal standards.

[3] In translation from Hebrew, we told our participants that: "The law requires that the plaintiff would show that his version is more probable than the defendant's. In what level should a judge be confident in the plaintiff's version to decide for the plaintiff? Express your answer in percentage terms from 0% (no need to believe that the plaintiff is right) to 100% (need to be completely convinced that the plaintiff is right)."

[4] For this group, we changed the first part of the relevant wording, *supra* note 3, in the following way: "The law requires that in order for the plaintiff to proceed beyond the preliminary phase, the plaintiff should present enough facts to state a claim for relief, which is plausible on its face."

[5] For this group, we changed the first part of the relevant wording, *supra* note 3, in the following way: " . . . to proceed beyond the preliminary phase, the plaintiff should present enough facts to state a claim for relief that does not seem to be frivolous."

[6] The online platform, Panel4All, *available at* https://www.panel4all.co.il, recruits paid participants that represent the general population (based on the following criteria: gender, age, religiousness, and district). We also guaranteed that, in each survey, our participants have not taken part in our previous experiments.

We would like to stress that the three legal standards we used are familiar in the Israeli legal system. The preponderance standard is the general standard under Israeli civil law. Likewise, the "plausible on its face" standard is quite similar to a pre-requisite to exempt indigent plaintiffs from filing fees. Court Rules (Court Fees), 2007, § 14(c) (Isr.). The "frivolous" criterion is used, for instance, as a consideration for fee-shifting. Rule 514 to the Rules of Civil Procedure, 1984, K.T. 5685, 2288 (Isr.).

These findings suggest the following. First, participants of the general population heed the differences between the two lower standards for dismissal and the preponderance of the evidence threshold (these differences are statistically significant). Second, the differences between the two versions of the standard for dismissal are indistinguishable. In the analysis that follows, then, we do not distinguish between these two standards for preliminary dismissal. Third, compared to the consensus among scholars and judges, our participants clearly exaggerated all legal thresholds. This result is in line with previous literature. In previous studies the preponderance rule, that in theory represents a 50% probability, is translated to ~70%-75% (Simon & Mahan 1971; Glöckner and Engel 2013, p. 242). While we are not familiar with experimental work on dismissal standards, we show that the standards to dismiss are also translated excessively—as they cannot be greater than the standard to win a case on the merits, which is thought to be 50% (Herman & Sacher 2015, p. 144). We note that we documented a statistically significant correlation between legal education and the interpretation of the standard—lawyers and law students translated the standards a bit lower than the general population, though the probabilistic translation of the legal standards by lawyers still seems excessive.[7]

Next, we present findings regarding the way in which participants report the legal standard *after* making a decision, and compare these numerical interpretations to our foregoing, no-decision figures. We start with results that pertain to the decision to dismiss at the outset; and proceed to results that pertain to the decision to accept the entire case.

3. Legal Thresholds and Dismissals Decisions

We presented other participants (N = 217) with a description of a legal case and asked them to decide whether to dismiss or not based on the legal threshold for dismissal. We used two vignettes—a tort case in which an old victim suffered a substantial personal injury in a nursing home, with questionable responsibility of the nursing home; [8] and a business

[7] Lawyers translated the preponderance of evidence standard to ~65%, and the two preliminary standards to ~50%.

[8] Specifically, the victim in this vignette suffered a severe deterioration to his (low) mental capacity. The family claimed that the victim fell from his bed at night, and they based their claim on the fact that the victim was rushed to a hospital to conduct an unscheduled computed tomography of his brain the day after the alleged fall.

dispute regarding the interpretation of a contract to purchase an office, with ambiguous wording.[9] We explained to the participants that a dismissal means that the plaintiff loses the case, and a decision not to dismiss means that in the next week, in the second part of the survey, they will have to decide the entire case based on additional evidence.[10] Immediately after their decision to dismiss (or not), we asked subjects to state, in probabilistic terms (0-100%), the strength of the plaintiff's claim and their interpretation of the legal rule. We also asked our participants to indicate, after their decision, the degree to which they are convinced in their decision, on a 1-5 scale.[11] Table 2 reports the combined results of these two experiments compared to the benchmark:[12]

Decision	Perceived legal standard	Perceived strength of the case	Confidence in the decision
Pro-plaintiff, not to dismiss (N = 189)	55% (27.4)	65% (19.9)	3.99 (0.78)
Pro-defendant, to dismiss (N = 28)	64% (26)	41% (28)	3.77 (0.75)
Benchmark (N = 183)	68% (24)		

Table 2. Means and standard deviations of the perceived legal standard (dismissals), merits of the case, and confidence in the decision (t1).

As the left column shows, most participants who had to decide whether to dismiss or not chose not to dismiss: $\frac{189}{189+28=217} \cong 87\%$. This figure seems plausible given the low standard for dismissal on the one hand and the questionable merits of both vignettes on the

[9] Specifically, the agreement included an obligation to provide a parking lot, and the plaintiff argued that the seller, the defendant/contractor, promised him a roofed parking lot. The contractor argued, in response, that the price the plaintiff paid for the parking lot reflects regular, rather than roofed, parking.

[10] To avoid incentives to skip the second part, we clarified that those who decide to dismiss will also have to fill a survey in the second part in order to get their reward. For a more detailed description see *infra* note 16.

[11] We also asked subjects to answer comprehension questions following the vignette (and before deciding), and eliminated participants who failed to show sufficient understanding—we filtered out those who did not answer correctly half or more of the comprehension questions, in both t1 an t2.

[12] As mentioned in the text, we used two different vignettes (a tort case and a business dispute). We also slightly changed the factual description in each vignette. The differences in the factual background between (and within) the two experiments are immaterial to the general trend that our findings depict. Therefore, we decided to collapse the reported results. As discussed before, we also collapsed the "plausibility" and the "non-frivolousness" standards for dismissal, as they were perceived as indistinguishable.

other hand. The next column demonstrates how the perception of the legal standard depends on the decision to dismiss (or not), relative to the benchmark, no-decision perception of the same standard. We conducted a univariate ANOVA with decision (dismiss / not-dismiss / no decision) as a between subjects variable. Decision significantly predicted perception of the legal standard ($F(2,396) = 12.376$, $p < .0001$, partial $\eta^2 = .06$). Planned contrasts showed that participants who did not dismiss differed significantly from participants in our benchmark (no decision) experiment ($\Psi = -13.26$, $p < .0001$, 95% CI [-18.52, -7.99]), while participants who dismissed did not ($\Psi = -4.23$, n.s).

The next column shows the perceived strength of the plaintiff's merits (on a 0-100% scale). Unsurprisingly, participants who decided for the defendant, i.e., to dismiss, perceived the case as considerably weaker than those who decided not to dismiss ($t(214) = 5.5$, $p = .0$, 95% CI [15.3, 32.2]). The last column shows the difference in the degree of confidence in one's decision, on a 1-5 scale. People who decided to dismiss seem less confident than those who decided not to dismiss, but these differences are not significant ($t(132) = 1.2$, $p = .23$, 95% Bootstrap CI [-.13, .56]).[13]

Table 3 reports the results of two OLS regressions in which the reported standard is the dependent variable, where we also controlled for one's perceived strength of the case and her merits decision (first model, second column), as well as the interaction between the perceived strength and the decision to dismiss (second model, third column). As Table 3 shows, there is a statistically significant correlation between our variables of interest, namely, the stated legal standard and the decision to dismiss:

Explanatory variables	(1) Coefficient (standard error)	(2) Coefficient (standard error)
Dismissal	19.9 (5.5)***	48.8 (10.5)***
Perceived strength	0.46 (0.1)***	0.6 (0.09)***
Interaction dismissal*strength		-0.63 (0.2)***

[13] One could expect to observe different confidence levels among those who dismissed and those who did not. A decision not to dismiss defer the final judgment to t2, while a decision to dismiss ends the case and seems harder to take. (We elaborate on this issue below, *infra* notes 16-17 and accompanying text).

Constant	24.8 (5.6)***	15.9 (6.2)**
Observations	216	216
R square	13.8%	17.8%

Table 3. OLS regressions with perceived standard as dependent variable. **, * indicate statistical significance at the 5% and 1% levels. The second model (third column) adds the interaction between the merits and the decision to dismiss to the first model (second column).**

These results suggest that the decision to dismiss was associated with a higher standard, while decision not to dismiss was associated with a lower standard. The differences in the perception of the dismissal standard between those who dismissed and those who did not are also non-trivial in size.

The differences between the dismissing and the non-dismissing participants' interpretations of the legal standard can stem from other, intervening factors, unrelated to the decision itself. It might be, for instance, that those who innately tend to interpret the standard relatively higher are also, by and large, pro-defendant; hence, we observe among the dismissing a higher reported standard, regardless of any shift in their perception of the standard. In a similar vein, the perception of the standard might move the decision to dismiss, and not the other way around, i.e., reverse causality.

To the extent that one's perceived strength of the case predicts the decision to dismiss and is also correlated with innate tendencies to interpret the standard differently, the regressions in Table 3, which control for one's perceived merits, suggest that the decision to dismiss has a unique contribution to the interpretation of the legal standard. To make a more robust case for the shift in the legal standard following one's decision, we separately ran hierarchical regressions in which we added to the foregoing explanatory variables several demographic characteristics, such as age, gender, marital status, education, and religion. In all our models the coefficient on the decision to dismiss remained positively correlated with the dependent variable, i.e., the perception of the legal standard. The decision to dismiss significantly predicted the perception of the legal standard over and above demographic characteristics.[14] While we cannot rule out alternative explanations,

[14] We did find several statistically significant (or marginally significant) correlations between demographic characteristics and the interpreted standard in the benchmark group (N=183)—age is positively correlated with the stated standard, while gender (men) and higher education were negatively correlated. In the dismissing/not-dismissing sample (N=217), and after adding the variables that indicate dismissing, perceived merits, and confidence, the demographic variables are no longer significant.

and in particular the reverse causation explanation, to the plausible extent that demographic variables move both the decision to dismiss and the interpretation of the standard, controlling for these characteristics better isolates the effect of dismissal on the interpretation of the standard.[15]

We conclude, then, that our t1 findings indicate a coherence-shift with regard to the legal standard, at least among those who decided not to dismiss—this group reported a lower standard, in line with its pro-plaintiff decision on the merits. We highlight the special context of this finding—a decision to dismiss at the outset (along the lines of *Twombly*), with rudimentary factual information. We caution that we found a shift in the reported standard among those who decided for the plaintiff, but not for those who decided to dismiss. Those who decided against the plaintiff, to dismiss, thought that the plaintiff's case was weaker, but, at least compared to the benchmark, no-decision group, they did not shift the standard in the direction of their decision (i.e., upward). This asymmetry could be explained by the fact that a decision not to dismiss is an intermediate decision (and was likewise framed in our surveys[16]); while a decision to dismiss ends the dispute.[17] We also stress that our findings are based on self-reports of the standard (Section 5 below discusses this point in greater detail).

The literature has studied the effect of legal thresholds on decision-making (e.g., Glöckner and Engel 2013; Engel 2009, p. 463); but less attention has been given to the effect of decision-making on legal thresholds. In particular, there is little discussion in the

[15] As a side note, we think that it is more plausible to believe that decision affects the stated standards in our experiments and not the other way around. First, the participants reported their perception of the standard immediately after deciding the case, suggesting the direction for causation. Second, the demographic characteristics seem to predict the decision on the merits better than the stated standard.

[16] At the dismissal stage our instructions read as follows (translated from Hebrew, emphasis added): "The defendant asks to dismiss the case at the outset. You are the judge assigned to the case, and you have to decide now whether the case should be dismissed at the outset. Dismissals occur in preliminary stages. *If you decide not to dismiss now, you will be presented next week with additional information, including relevant testimonies, with which you will decide the case. If you decide to dismiss the case at this stage, the plaintiff loses the entire case*. Whether you decide to dismiss or not, you will be asked to participate next week in the second part of this survey, after which you will receive your reward."

[17] There are several other possible conjectures for this asymmetry. The proposition that those who dismissed did not shift the standard relate to the comparison to those who interpreted the standard in the abstract, no-decision condition. To the extent participants particularly exaggerate the standard in the abstract, such a comparison might mask an upward shift among those who dismissed. Relatedly, the decision to dismiss seems to be an unusual decision, as it eliminates access to court in preliminary stages (and only 28 participants decided to dismiss). These characteristics might render the numerical assessment of the dismissal standard particularly difficult in the abstract.

literature regarding coherence-based shifts of legal standards. The most relevant paper is Phillips (2002). In Phillips's research, participants assumed the role of an accountant and had to decide whether a company survives the "significant doubt" threshold concerning its financial situation. In addition, participants had to rate, on a 0-100% scale, their interpretation of this accounting rule. It was found that participants distorted the numerical interpretation of the accounting threshold in the direction of their decision. Those who decided for the company reported that the threshold translates to ~53%, whereas those who decided against the company interpreted the threshold as more onerous, ~62%.[18] However, other than Phillips (2002), we found no other studies that replicated such a shift in the interpretation of legal standards (compare Scurich 2012, pp. 61-64, 82-84).[19]

4. Legal Thresholds and Deciding the Entire Case

a. General findings—no coherence shift at t2

In the second part of our experiments we asked participants to decide the entire case and report their interpretation of the applicable standard. As discussed in Section 3, our subjects received some information at the first stage, t1 (where some participants had to decide at this stage whether to dismiss or not, on a lower standard—we later discuss this variation). One week after, at t2, they received additional information with instructions to accept or reject the plaintiff's claim based on the preponderance, "more probable than not" threshold.[20] As before, immediately after deciding we asked the participants to rate in

[18] More precisely, Phillips (2002) manipulated the timing of the question to rate the standard—before or after deciding "on the merits," and the numbers in the text refer to the post-decision interpretation of the standard. Our research examines the post-decision interpretation of the standards, and compares it to a benchmark, no-decision group, that the standard in the abstract.

[19] Another relevant study is Glöckner and Engel, who show experimentally that those who decided to convict in a criminal case reported a lower standard, in numerical terms, than those who decided to acquit (Glöckner and Engel 2013, pp. 242-43). The change is ~5.4 percentage points. Glöckner and Engel did not compare the post-decision perceptions of the standards to a benchmark, no-decision group, and did not control for the correlation between the subjective strength of the case and the participants' demographic characteristics to the reported interpretation of the standard.

In another relevant research, Scurich finds evidence of a small shift in the legal threshold following a decision. This work, though, studies the implicit threshold, inferred from participants' decision on the merits and their responses regarding the strength of the case, rather than the explicit, reported numerical interpretation of the threshold (Scurich 2012, pp. 68-106).

[20] As before, we used two vignettes with some modifications within the vignettes. *Supra* notes 8-9, 12 and accompanying text. For the preliminary information given to the participants at the first stage see *supra* notes 8-9. In the torts case, the additional information at the second stage included the following: the plaintiff

probabilistic terms the strength of the plaintiff's claim and their interpretation of the legal threshold (both on a 0-100% scale), as well as their confidence in their decision (on a 1-5 scale). Table 4 summarizes the results (N = 351):

Decision	Perceived legal standard	Perceived strength	Confidence in the decision
Accept (N = 236)	79% (19.1)	76% (17.2)	3.67 (0.7)
Reject (N = 115)	73% (23.2)	35% (24.7)	3.74 (0.7)
Benchmark (N = 183)	81% (21.9)		

Table 4. Means and standard deviations of the perceived legal standard (final decision), merits of the case, and confidence in the decision (t2).

The left column shows that most participants accepted the plaintiff's case, $\frac{236}{351} \cong 67\%$. Unsurprisingly, this rate is lower than the rate at which participants ruled for the plaintiff at the dismissal stage, under a lower standard.[21] We focus on the first column, which demonstrates how the perception of the "more probable than not" standard changes per one's decision, relative to the benchmark, no-decision group.[22] One can observe that the pro-plaintiff, accepters' reported standard (79%) is close to the benchmark group (81%), indicating no coherence shift. Surprisingly, those who decided to reject reported a lower

presents the opinion of a medical expert who indicated that the plaintiff's mental deterioration was sudden—which may have stemmed from a fall. The victim's family also presents testimonies regarding the victim's tendency to wake up at night, which should allegedly have prompted the nursing home to take extra care. The nursing home's staff conceded that the victim suffered from a severe mental deterioration, and that he was taken care of according to the regular procedures (and no extra care was taken). Nonetheless, the nurses asserted that no special events occurred; that the victim's deterioration results from natural fluctuations in his mental condition; and that he was taken to a computed tomography of his brain as part of a routine, but non-scheduled check up.

In the business dispute, the additional information included the following: the plaintiff's partner testifies that the contractor promised the plaintiff that the parking lot would be in the building, i.e., a roofed parking lot. The plaintiff provides evidence that he paid for the parking in cash, hence, allegedly he received a discount, paying less than the usual price for a roofed parking. The contractor concedes that he offered a parking in the building, but he asserts that he indicated that the parking lot will be at the top of the building, i.e., unroofed. The contractor's employees present evidence that suggests that no other tenant received a discount on the price of a roofed parking, even for paying cash, and that the roofed parking has a fixed price (above the sum that the plaintiff paid).

[21] We attempted to design the additional information at t2 (*supra* note 20) as neutral. Hence, the difference in the proportion of pro-plaintiff decisions at t1 and t2 seems to result from the more onerous standard at t2, indicating that our subjects generally managed to implement different standards.

[22] The second column shows, in essence, that those who rejected the case thought that it is weaker (35%) than those who accepted (76%); this difference is statistically significant ($t(349) = 18.0$). The third column shows that people who rejected ($M = 3.74$, $SD = 0.7$) or accepted ($M = 3.67$, $SD = 0.7$) the case did not differ in a statistically significant way with respect to their decision confidence ($t(349) = -0.81$, n.s.).

perception (73%) of the standard. To verify these results we conducted a univariate ANOVA with decision (reject / accept / no required decision) as a between subjects variable. Decision significantly predicted perception of the legal standard ($F(2,531) = 4.33$, $p = .01$, partial $\eta^2 = .02$). Planned contrasts showed that participants who rejected the case differed significantly from participants in our benchmark (no decision) experiment ($\Psi = -7.33$, $p = .004$, 95% CI [-12.24, -2.41]), while participants who accepted the case did not ($\Psi = -2.28$, n.s). The tendency of rejectors to lower the standard is counter-intuitive—on its face, by lowering the standard they made their case to reject harder, contrary to the coherence-based model. We refer to this phenomenon as a "reverse-shift" in the perception of the legal standard, i.e., against one's own decision, and discuss it in further detail in the next sub-section.

We conclude that the findings in Table 4 indicate no coherence-shift of the legal standard at t2. They are inconsistent, then, with the views that equate legal standards with evidence. Interestingly, unlike the perception of the standard, the perception of the *evidence* did shift in the decision's direction. Table 5 summarizes our findings along these lines and demonstrates how subjects change their perception of the legal standard and the evidence at t1 (upper panel) and t2 (lower panel):[23]

Timing	Decision	Perceived strength	T test	Perceived legal standard	T test
Decision at t1 (dismissal)	Pro-plaintiff, not to dismiss (N=189)	65%	5.5***	55%	-1.7*
	Pro-defendant, to dismiss (N=28)	41%		64%	
Decision at t2 (final)	Pro-plaintiff, to accept (N=236)	76%	18.0***	79%	2.2**
	Prof-defendant, to reject (N=115)	35%		73%	

Table 5. Perceptions (means) of the evidence and the legal standard, at t1 (upper panel) and t2 (lower panel). The columns T test show the t-test values of the differences between pro-defendant and pro-plaintiff at each cell—where *,, *** indicate statistical significance at the 10%, 5%, and 1% levels.**

[23] In Figure 1 in the Introduction, we demonstrated the same data graphically.

While at t1 pro-plaintiff decisions-makers rated the strength of the case as 65% and pro-defendant subjects rated it as 41%, at t2 participants' views on the case polarized considerably. Pro-defendant decision-makers rated the case as 35% and pro-plaintiff subjects rated the case's strength at 76%. This is, of course, consistent with the coherence-based model of reasoning.[24] However, we observe no such polarization with respect to the legal standard at t2. At least at the final stages, then, with more information, participants do not react to the standard as another piece of evidence.

b. Reverse-shift and a conflicting decision at t1

The foregoing demonstrates that participants did not shift the perception of the legal standard to cohere with their *final* decision. In fact, Tables 4 and 5 suggest that, if anything, participants shifted the standard in the reverse direction, i.e., opposite to their decision. In the abstract, participants in the benchmark group interpreted the preponderance standard as ~81%; those who accepted the plaintiff's claim at t2 reported a close standard (78%); but those who rejected (73%) stated a lower standard. Why would pro-defendant decision-makers choose to lower the standard, making the plaintiff's case easier?

Previously we concluded that participants at t2 treated the standard separately than other pieces of evidence. We suggest that this distinct consideration of the legal standard, together with a reverse-shift in its perception, could actually serve to justify the strength of the decision-maker's position. To illustrate, the pro-defendant decision-maker might believe that the evidence for the defendant is sufficiently strong such that it warrants a decision for the defendant, even under a lower standard of proof. In particular, our logic resembles "self-signaling." It was suggested that in noisy environments individuals engage in costly behaviors—such as volunteering—as a way to signal personal traits to themselves (Bénabou & Tirole 2006). Along these lines, those who decide a case with a given set of evidence can better justify their decision through a reverse-shift in the applicable standard. This use of the legal standard, to self-signal and justify one's decision, is consistent with the two-step, legalistic approach in which the legal standards is taken into account separately from the evidence (and after one has processed the evidence).

[24] We note here that the additional information at t2 was designed to be neutral, with an equal number of evidence that support each side. *Supra* notes 16 and 21.

We further diced the data in order to better understand the tendency to reverse-shift the standard of proof. Recall that our data includes participants who were asked to decide at t1 to dismiss, per a lower standard and with preliminary information ("treatment group"); and participants who were asked to decide only the entire case, at t2 ("control group"). The following table shows that the tendency to lower the standard stems from the treatment group:

Decision	Overall	T test	Control groups	T test	Treatment groups	T test
Accept	78.5%	2.2**	78.6% (23.0) (N=84)	0.7	78.5% (16.7) (N=152)	2.5**
Reject	73.5%		75.6% (25.3) (N=50)		71.8% (21.5) (N=65)	
Benchmark (N = 183)	81% (21.9)					

Table 6. Means and standard deviations of the perceived legal standard (final decision), by treatment and control groups. The columns T test show the t-test values of the differences between pro-defendant and pro-plaintiff at each cell—where ** , * indicate statistical significance at the 5% and 1% levels.**

As Table 6 shows, those who rejected the case, both in the treatment and control groups, reported a lower interpretation of the preponderance standard than those who accepted. However, the reduction in the reported standard among rejectors in the control group is smaller (and statistically insignificant) than the reduction in the standard among rejectors in the treatment group. Why would subjects in the treatment group—who were forced to decide the case at t1—shift their interpretation of the standard? The following, Table 7, breaks the rejectors in the treatment group to those who decided to dismiss at t1 and those who decided not to dismiss at t1:

Decision	Perceived legal standard at t2	T test	Perceived strength at t2	T test	Confidence in the decision at t2	T test
Dismissed at t1 and rejected at t2 (N=16)	79% (22.5)	-1.6	36% (21.5)	-0.3	3.81 (0.7)	-0.4

Did not dismiss at t1 but rejected at t2 (N=48)	69% (21.0)	34% (24.0)	3.73 (0.6)
Benchmark (N = 183)	81%		

Table 7. Means and standard deviations of the perceived legal standard (final decision), perceived strength and confidence (all at t2), for participants in the treatment group who rejected at t2, by dismissed and not-dismissed at t1. The columns T test show the t-test values of the differences between pro-defendant and pro-plaintiff (at t1) at each cell.

Table 7 shows that those who decided first that the case survives dismissal (for the plaintiff), but then with more information rejected the final case, lowered their interpretation of the preponderance standard, in an opposite direction to their decision. However, those who dismissed at t1 and likewise decided against the plaintiff at t2 reported a standard close to the benchmark, no-decision one. The reverse-shift in the standard, then, appears to stem from those who made a conflicting decision—for the plaintiff at t1 but against the plaintiff at t2. Those participants were seemingly under the greatest need in justifying their decision at t2. They eventually decided to deny the plaintiff's case, but nonetheless reported a lower standard to win, as if they "signal" to themselves their conviction in their own decision. One alternative explanation to this proposition is that those who ruled for the plaintiff at t1 (but then rejected), have an independent inclination to state a lower standard (indeed, they ruled for the plaintiff at t1, suggesting that their stated standard at t1 was lower). Although we cannot completely rule out this alternative explanation for the reverse-shift we documented, the findings that we presented in Tables 6 and 7 seem to fit our description, i.e., that our participants self-signal due to their conflicting decision at t1.[25]

Participants who chose to reject the case reported a lower interpretation of the preponderance standard than those who chose to accept (and also lower than those in the

[25] We do not think that the foregoing alternative explanation fully describes the lower standards reported by those who had a previous, conflicting decision at t1. First, as Table 6 suggests, rejectors at t2 reported overall a lower standard when they were forced to take a decision at t1, i.e., in the treatment group. Second, as Table 7 suggests, among the rejectors at t2 in the treatment group, those who dismissed at t1 and those who did not dismiss at t1 have similar views on the merits of the case (36% and 34%, respectively). Nonetheless, the rejectors who did not dismiss at t1 reported a lower standard. This plausibly reflects a similar perception of the evidence, but a self-justificatory approach to the standard.

benchmark group, who interpreted this legal threshold in the abstract, without deciding). We refer to this finding as a "reverse-shift" in the standard, and suggest self-signaling as an explanation for this surprising phenomenon. Whether it stems from self-signaling or not, the general results concerning reverse-shift (Table 6) are inconsistent with the coherence-based model of reasoning. Further research would be able to better identify the sources of such inverse shift in the interpretation of the standard.

* * *

We ran several experiments in which we asked subjects to decide a case and report their numerical interpretation of the standard. We sketch here our main findings. Our first set of findings relate to the exaggerated numerical interpretation of legal thresholds. Consistent with previous literature, we find that subjects tend to exaggerate the numerical interpretation of the well-known preponderance of the evidence standard. In addition to the preponderance rule, we asked subjects to rate dismissal, *Twombly*-like standards. We found that subjects also translate dismissal standards as higher than they should be. In this respect, we also found that the "plausibility" and the "non-frivolousness" standards for dismissal are indistinguishable. While previous experimental literature has not examined dismissal standards, the results in our study are in line with the general tendency to exaggerate the numerical interpretation of legal standards.

The remainder of our findings pertains to the "shift" in legal standards, depending on one's merits decision. First, participants who were asked to take an intermediate decision—whether to dismiss a case with rudimentary information (at t1)—and decided not to dismiss, seemingly shifted the standard in the direction that fits their decision. This outcome is consistent with the coherence-based logic, indicating that subjects treated the standard as other pieces of evidence. However, we did not find a similar tendency to shift the standard in other settings. In particular, our participants did not shift the standard in the direction of their decision when they had to take a final decision, at t2. We thus conjecture that in the context of intermediate decision-making, with limited information, subjects could treat the standard similarly to other evidence. Alternatively put, in these settings legal thresholds are decision-forming. However, in general legal thresholds are treated differently than

other pieces of evidence, as the two-step, legalistic view predicts. We have found scarce evidence in the literature regarding coherence-shifts in the perception of thresholds. As we found coherence-shift in legal thresholds only in limited circumstances, namely, in an intermediate decision, our results can explain this scarcity.

Second, counter-intuitively, some of the subjects at t2 interpreted the legal threshold in a reverse direction, i.e., opposite to their merits decision at t2. We conjecture that the tendency to reverse-shift the legal standard serves as a decision-justifying mechanism, resembling "self-signaling." Indeed, the reverse-shift findings mainly stem from a group of subjects who decided a conflicting decision (for the plaintiff) at t1, and hence were under the greatest need in justifying their decision at t2.

Our overarching conclusion is that legal standards are treated materially different from other evidence. Except for non-dismissing at the intermediate, t1 decision to dismiss—a unique context—we observe no coherence-shift in the legal standard. This conclusion corroborates the legalistic description of the legal standard—an independent cognitive operator in a two-step decision process. It implies that the consideration of the legal standard mitigates the repercussions of coherence-based decision-making, along the general lines of Glöckner and Engel (2013) and Engel (2009, pp. 457-465). We should expect, then, that the explicit inclusion of the legal standard in a decision-making task would lead to less polarization and confident in one's overall decision. In line with this prediction, we remark here that our subjects, both pro- and anti-plaintiff, reported less confidence in their decision at t2 than at t1,[26] although their perception of the merits polarized from t1 to t2 (see Figure 1 and Table 5). Future research would be able to better identify the subtle effects of legal standards on decision-making.

In the next, concluding section, we discuss the broader implications of the interdependence between one's decision and the stated standard, and demonstrate their relevance to actual legal decision-making.

[26] At t1, those who decided for (against) the plaintiff reported a confidence level of 3.99 (3.77). At t2, those who decided for (against) the plaintiff reported a confidence level of 3.67 (3.74). See, respectively, Tables 2 and 4.

5. Broader Discussion and Implications

Our findings demonstrate how legal standards could be interpreted differently, depending on one's substantive decision in the case. We note again that our findings relate to the reported standards, rather than their actual application. Nonetheless, to the extent decision-makers bias the reported standard, this tendency could have various legal implications.

We illustrate this point through a brief discussion of the recent developments in the dismissal standards in the federal courts in the U.S. Attempting to cope with the problem of meritless suits and the soaring costs of discovery, the U.S. Supreme Court decided in *Bell Atlantic Corp. v. Twombly* (550 U.S. 544 [2007]) to demand more substance from plaintiffs who file complaints (also, Ashcroft v. Iqbal, 556 U.S. 662 [2009]). In order to survive dismissal motions and proceed to discovery, plaintiff are now required to show "enough facts to state a claim to relief that is plausible on its face." Beforehand, the standard to proceed to discovery was considered as considerably lower— "a short and plain statement of the claim," giving the defendant "fair notice of what the plaintiff's claim is" (*Conley v. Gibson*, 355 U.S. 41 [1957], p. 47).[27] The Supreme Court offered no meaningful guidance concerning the application of the new "plausibility" legal threshold, and lower courts have struggled to interpret it (Reinert 2012). Nonetheless, it is widely considered that the new dismissal standard constitutes a major doctrinal shift, designed to make it harder for plaintiffs to survive dismissals (Hubbard 2016, pp. 695-696). Indeed, the doctrinal change was explicitly intended to weed out unwanted cases and curb the discovery costs of defendants. Surprisingly, however, it seems that at best little real-world change occurred. The relevant empirical literature on this question appears to be ambiguous. Analyzing the empirical literature, for instance, one commentator concludes "that plausibility pleading has wrought no major change in practice" (Hubbard 2016, p. 699).

Consider in this respect the findings of this paper. As even the post-*Twombly* dismissal standard is not too onerous, it seems that most cases should survive dismissal. Along these

[27] The previous precedent also directed "that a complaint should not be dismissed . . . unless it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim . . . " (*Conley v. Gibson*, 355 U.S. 41 [1957], pp. 45-46).

lines, in our surveys ~87% of the subjects chose not to dismiss. However, those who decided not to dismiss shifted the standard, reporting a relatively low threshold. To the extent such coherence-shifts can be extended to real-world judging, decisions not to dismiss (apparently, most decisions) would state a relatively low dismissal standard. As the Supreme Court left the interpretation of the standard to lower courts, subsequent judges who rely on past decisions of lower courts would again choose not to dismiss in most cases, again reporting an even lower interpretation of the dismissal standard. This process shows how coherence-based shifts, similar to those that we observed in our experiments, could degrade the effectiveness of legal thresholds over time. This process might be one reason for which the actual application of the *Twombly*-style dismissal standard in the U.S. seems inconsequential. Of course, one can think of various complications to this process.[28] But more generally, a tendency to adjust legal standards risks biasing future decisions, and it could be particularly salient in the context of intermediate, preliminary decisions to dismiss. As a side note, our findings also show that pro-defendant decision-makers, who decided not to dismiss at t1 but later ruled against the plaintiff, reported a weaker interpretation of the preponderance standard (what we referred to as "reverse-shift" of the standard). Hence, the introduction of stricter dismissal standards could further bias the interpretation of legal standards post-*Twombly* in favor of plaintiffs.

It is, of course, not trivial to extend experimental results, which are based on laypeople's responses, to the real world in general and to actual judicial behavior in particular. Indeed, in the real-world people have the ability to somewhat correct, over time, mistakes that have been observed in a one-shot experiment (Kuilen & Wakker 2006). On the other hand, previous literature has shown that judges are vulnerable to similar psychological manipulations as the general population (Rachlinksi et al. 2013). Moreover, while learning is possible, it is typically facilitated by feedback (Wilson et al. 2002; Kuilen & Wakker 2006). Such a feedback seems uncommon in the context of trial court adjudication. Judges, then, could be vulnerable to the modes of reasoning that we described in this paper. Specifically, it is not implausible to believe that judges rely on previous

[28] For instance, this description has not taken into account the behavior of defendants, and in particular repeat-defendants, who have to trigger dismissal motions under the current regime. Anecdotally, surveys suggest that defendants are not inclined to move to dismiss after *Twombly* (Hubbard 2016, p. 737).

judges' interpretations, and that they report a biased interpretation of vague legal maxims, depending on their decision on the merits (Simon 2004, p. 549).

Be that as it may, little literature discusses and examines the interplay between one's decision and her stated legal standard. As we demonstrated, the relations between these two factors have potentially wide implications, especially where judges rely on previous judges' interpretation of the law. At the least, our results suggest that the interaction between stated legal thresholds and merits decisions is not trivial. Future literature will be able to shed more light on this question.

Bibliography

- Ronald Bénabou & Jean Tirole, *Incentives and Prosocial Behavior*, 96 AM. ECON. REV. 1652 (2006).
- Robert G. Bone, *Twombly, Pleading Rules, and the Regulation of Court Access*, 94 IOWA L. REV. 873 (2009)
- Kevin M. Clermont, *Standards of Proof Revisited*, 33 VT. L. REV. 469 (2009)
- Christoph Engel, *Preponderance of the Evidence Versus Intime Conviction: A Behavioral Perspective on a Conflict Between American and Continental European Law*, 33 VT. L. REV. 435 (2009)
- David Freeman Engstrom, *The* Twiqbal *Puzzle and Empirical Study of Civil Procedure*, 65 STAN. L. REV. 1203 (2013)
- Vincy Fon & Francesco Parisi, *On the Optimal Specificity of Legal Rules*, 3 J. INSTITUTIONAL ECON. 147 (2007)
- Andreas Glöckner and Christoph Engel, *Can We Trust Intuitive Jurors? Standards of Proof and the Probative Value of Evidence in Coherence-Based Reasoning*, 10 J. EMPIR. LEGAL STUD. 230 (2013)
- Alice Guerra et al., Standards of Proof and Civil Litigation: A Game-Theoretic Analysis, B.E. J. Theoretical Econ. (forthcoming 2018)
- Douglas A. Herman & Seth B. Sacher, *An Economic Analysis of* Twombly/Iqbal *with Applications to Antitrust*, 11 J. COMP. L. & ECON. 107 (2015)
- William H.J. Hubbard, *A Fresh Look at Plausibility Pleading*, 83 U. CHI. L. REV. 693 (2016)
- Jason Scott Johnston, *Uncertainty, Chaos, and The Torts Process: An Economic Analysis of Legal Form*, 76 CORNELL L. REV. 341 (1991)
- Dorothy K. Kagehiro, *Defining the Standard of Proof in Jury Instructions*, 1 Psychological Science 194 (1990)
- Dorothy K. Kagehiro & W. Clark Stanton, *Legal vs. Quantified Definitions of Standards of Proof*, 9 LAW & HUMAN BEHAVIOR 159 (1985)
- Louis Kaplow, *Rules versus Standards: An Economic Approach*, 42 DUKE L. J. 557 (1992)
- Louis Kaplow, *Burden of Proof*, 121 YALE L. J. 738 (2012)

- Gijs van de Kuilen & Peter P. Wakker, *Learning in the Allais Paradox*, 33 J. RISK & UNCERTAINTY 155 (2006)

- Ziva Kunda, *The Case for Motivated Reasoning*, 108 Psychological Bulletin, 480 (1990)

- Michael S. Pardo, *Group Agency and Legal Proof; or, Why the Jury Is an "It,"* 56 WM. & MARY L. REV. 1793 (2015)

- Fred Phillips. *The Distortion of Criteria after Decision-making*, 88 ORG. BEHAV. & HUM. DECISION PROCESSES 769 (2002).

- NICHOLAS SCURICH, THE DYNAMICS OF REASONABLE DOUBT (Ph.D. Dissertation, on file with authors).

- Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 UNIV. CHICAGO L. REV. 511 (2004).

- Jeffrey J. Rachlinski, Andrew J. Wistrich, Chris Guthrie, *Altering Attention in Adjudication*, 60 UCLA L REV. 1586 (2013).

- Jeffrey J. Rachlinski, *Processing Pleadings and the Psychology of Prejudgment*, 60 DEPAUL L. REV. 413 (2011)

- Alex Reinert, *Pleading As Information-Forcing*, 75 LAW & CONTEMP. PROBS., 1 (2012).

- Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 Law & Soc'y Rev. 319 (1971).

- Fredrick E. Vars, *Toward a General Theory of Standards of Proof*, 60 CATH. U. L. REV. 1 (2010)

- Timothy D. Wilson et al., *Mental Contamination and the Debiasing Problem*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 185 (Thomas Gilovich et al., eds., 2002)

- William Douglas Woody & Edie Greene, (2012), *Jurors' Use of Standards of Proof in Decisions about Punitive Damages*, 30 BEHAV. SCI. & LAW 856 (2012)

- William Douglas Woody, Ph.D.* and Edie Greene, Ph.D.†

- Eyal Zamir and Ilana Ritov, *Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation*, 41 J. LEGAL STUD. 165 (2012).