UC Berkeley
UC Berkeley Previously Published Works

Title

From scanner to court: A neuroscientifically informed "reasonable person" test of trademark infringement

Permalink

https://escholarship.org/uc/item/28x3j016

Journal

Science Advances, 9(6)

ISSN

2375-2548

Authors

Zhang, Zhihao

Good, Maxwell

Kulikov, Vera

et al.

Publication Date

2023-02-10

DOI

10.1126/sciadv.abo1095

Copyright Information

Peer reviewed

COGNITIVE NEUROSCIENCE

From scanner to court: A neuroscientifically informed "reasonable person" test of trademark infringement

Zhihao Zhang[1,2,3]*, Maxwell Good[3,4,5], Vera Kulikov[3], Femke van Horen[6], Mark Bartholomew[7], Andrew S. Kayser[4,5,8]†* , Ming Hsu[3,4]†*

Many legal decisions center on the thoughts or perceptions of some idealized group of individuals, referred to variously as the "average person," "the typical consumer," or the "reasonable person." Substantial concerns exist, however, regarding the subjectivity and vulnerability to biases inherent in conventional means of assessing such responses, particularly the use of self-report evidence. Here, we addressed these concerns by complementing self-report evidence with neural data to inform the mental representations in question. Using an example from intellectual property law, we demonstrate that it is possible to construct a parsimonious neural index of visual similarity that can inform the reasonable person test of trademark infringement. Moreover, when aggregated across multiple participants, this index was able to detect experimenter-induced biases in self-report surveys in a sensitive and replicable fashion. Together, these findings potentially broaden the possibilities for neuroscientific data to inform legal decision-making across a range of settings.

INTRODUCTION

Did the song "Blurred Lines" plagiarize Marvin Gaye's "Got to Give it Up" (*1*)? Does the toothpaste Colddate infringe upon the trademark of Colgate (*2*)? In these and many other cases, a set of important legal questions centers on the thoughts or perceptions of some idealized group of individuals, variously conceptualized as the "average person," the "typical consumer," or simply the "reasonable person" (*3*–*8*).

Despite the seemingly commonsensical nature of such questions, their legal resolutions are often criticized for a perceived vulnerability to bias and manipulation (*6*, *7*, *9*, *10*). Even rigorous survey methods, currently seen as the most scientifically valid approach, face considerable skepticism (*9*–*12*). First, because responding to a survey item is itself a complex cognitive process, survey responses are known to exhibit context-dependent effects because of question wording, item order, response options, and other factors. Second, attempts to demonstrate bias in self-report instruments, either its presence or absence, are known to be arduous, contentious, and very often inconclusive. Thus, even when the presence of bias is virtually certain, such as when mutually contradictory findings are presented, the court may nevertheless still find itself unable to establish the relative credibility of the respective pieces of evidence. Perhaps as expected, judges have been known to forgo external evidence when dealing with these questions and instead to rely on their own beliefs and predispositions (*10*, *12*).

Here, we propose that neuroscientific data can provide an effective means of improving the state of evidence-based legal decision-making in the class of cases that center on the thoughts or perceptions of some idealized group of individuals. Specifically, we sought to assess the possible evidentiary value of neuroscientific data by focusing on disputes involving arguably the best understood set of processes in modern neuroscience—those involved in visual processing. To this end, we investigated a class of intellectual property law that evaluates whether a trademark is so similar to another as to generate consumer confusion (*3*, *9*, *13*, *14*). Because visual cues, such as trademarks and package designs, play an outsized role in determining how consumers respond to products, laws governing trademarks forbid counterfeit and look-alike products on grounds that they harm consumers by misleading potential buyers (*13*, *14*).

In disputes involving trademark infringement, plaintiffs must therefore show that the alleged infringing design causes consumer confusion. A typical recitation of the factors that contribute to confusion asks the trier of fact to assess "the strength of the senior user's mark, the degree of similarity between the two marks, the proximity of the products, the likelihood that the prior owner will bridge the gap, actual confusion, the junior user's good faith in adopting its own mark, the quality of the defendant's product, and the sophistication of the buyers" (*15*). While the number of such criteria that legally describes "consumer confusion" varies depending on the jurisdiction, our focus on visual similarity is motivated by findings from empirical studies of legal decision-making showing that assessments of "visual similarity" exert by far the greatest weight on the court's judgment (*10*, *16*).

Our focus on representative assessments of visual similarity offers several important advantages in maximizing the utility of neuroscientific data to the law. First, because the question of interest does not hinge on the mental state of a specific individual, we can sidestep what is sometimes called the group to individual ("G2i") inference problem in scientific expert testimony (*17*, *18*). This problem refers to the difficulty of translating scientific findings regarding general mechanisms, which are typically established on an aggregate level (e.g., factors that influence accuracy of eyewitness

[1]Darden School of Business, University of Virginia, Charlottesville, VA, USA. [2]Department of Psychology, University of Virginia, Charlottesville, VA, USA. [3]Haas School of Business, University of California, Berkeley, Berkeley, CA, USA. [4]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [5]Department of Veterans Affairs Northern California Health Care System, CA, USA. [6]School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [7]School of Law, State University of New York at Buffalo, Buffalo, NY, USA. [8]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA.
†These authors contributed equally to this work.
*Corresponding author. Email: zhangz@darden.virginia.edu (Z.Z.); andrew.kayser@ucsf.edu (A.S.K.); mhsu@haas.berkeley.edu (M.H.)

recollections), to address questions pertaining to a specific individual in a particular case at hand (e.g., the accuracy of a particular eyewitness's testimony). In contrast, the intrinsic reliance of representativeness on aggregate responses reduces the demand on the precision of neuroscientific methods, where limited signal-to-noise ratio and spatiotemporal precision complicate interpretation in single subjects (11, 18).

Second, by focusing on questions of visual similarity, we leverage current knowledge of the visual system (19), which can be seen as providing an "upper bound" on the discriminatory power of neural data. Substantial evidence, for example, indicates that regions within the fusiform and inferotemporal cortices engage in holistic, as opposed to parts-based, representation of objects (20–23). Moreover, the deep history of experimental studies that produced this knowledge offers a number of analytical approaches to capture different aspects of visual representations, such as repetition suppression (RS), which provides a readout of similarity between two stimuli without requiring additional assumptions about how to quantify similarity between representations (24–27). Last, despite skepticism surrounding the use of self-report data, the fact that they can be accepted by the court offers an opportunity for neuroscientific data to either buttress or challenge their validity (28). Trademark infringement cases, which routinely include survey evidence even if it is, at times, discounted by the court (10, 11), allow us to design tests capable of addressing concerns of bias.

To demonstrate the added value of neuroscientific measurements, here we tested the extent to which a neural index of visual similarity based on blood oxygen level–dependent (BOLD) signals measured by functional magnetic resonance imaging (fMRI) is capable of detecting biases in self-report instruments (Fig. 1). Using formats and language commonly used in trademark cases, we constructed a set of surveys that included experimenter-induced biases that drew upon past criticisms of litigation surveys (9, 29). Specifically, we manipulated the survey design such that results varied in how strongly they favored either the plaintiff or the defendant. We found that our neural index was sufficiently precise to detect examples of these experimenter-induced biases, suggesting that combining neural and self-report measures may provide a more robust measure than either measure alone. We conclude by discussing possible ways to improve our proposed tool and broaden its domain of application.

RESULTS
Creating realistic test stimuli
Given our goal of developing a realistic simulation of actual legal cases, we considered a scenario involving potential U.S. trademark infringement of a common candy, Reese's Peanut Butter Cups, as well as a common laundry detergent, OxiClean. Reese's was selected in part because of its role in a 2014 lawsuit to prevent the import of the British candy Toffee Crisp on infringement grounds (30). OxiClean was selected to create variation in visual appearance of the trademarks (e.g., color) and to evaluate a nonappetitive item.

For each category, either candy or cleaning product, we created a set of comparison products that varied in visual similarity assessed according to pretests (Supplementary Methods). The goal for including these comparison products was (i) to better demonstrate the extent to which biases in similarity judgments may be introduced by different self-report surveys and (ii) to better determine

the ability of the proposed neural index to distinguish between different levels of similarity. Some stimuli, such as Toffee Crisp or Tide, were based on real products, whereas others, such as Pieces and Breeze, were fictitious. Notably, the visual appearance of the packaging of Toffee Crisp (referred to as the "trade dress" in trademark law) was determined by the court in the lawsuit mentioned above to be confusingly similar to that of Reese's, leading its import to the United States to be discontinued (30). We also included two cases of real product variants that are from the same manufacturers and intended to be of high similarity to (yet not the same as) Reese's and OxiClean: a brand extension product of Reese's (Reese's Sticks) and an international version of OxiClean. Hereafter, we refer to Reese's and OxiClean as the "reference product" for their respective categories, whereas other products are described as "competitor products."

Development of an experimental test bed to observe and manipulate bias
Assessments of bias in self-report instruments are often challenging because of the lack of a gold standard (31). We sought to overcome this issue by creating a testing environment in which bias can be experimentally manipulated and calibrated in a transparent and replicable manner. Specifically, we developed a set of surveys in a hypothetical legal setting that contained varying degrees of bias. These biases were experimentally induced and designed to favor either proposed plaintiffs (Reese's and OxiClean) or potential defendants (Pieces and OxyClear).

Using formats and language commonly used in trademark cases, our surveys drew upon documented criticisms of litigation surveys presented in trademark infringement lawsuits (10, 32, 33) and the recent scientific literature on "questionable research practices" that greatly inflate the likelihood of false-positive findings (34). As our purpose is to produce systematic biases to achieve some preferred outcome, we did not attempt to closely match these different versions. Specifically, we induced biases through (i) explicit means, such as referring to trademark infringing products as "copycats" for the putatively Pro-Plaintiff survey and companies pursuing trademark infringement lawsuits as "trademark bullies" for the putatively Pro-Defendant survey, and (ii) more subtle means, such as the format of the questions, the criteria for making similarity judgments, and the scale of the response (e.g., the Likert scale or binary judgments) (table S1; see details in Supplementary Methods).

These design elements constitute a large space of features that survey creators need to navigate without agreed-upon standards. As a result, design decisions are often targets of contentions in court. While our specific ways of creating bias do not necessarily reflect those in real-world litigation surveys with perfect fidelity, they nevertheless provide a positive control to examine the ability of our method to detect known bias. Starting with more extreme forms of bias here also paves the way for testing our method with more moderate bias (see the "Replicability and sensitivity of RS index" section below).

Comparison between results from the three surveys based on data collected from independent samples on Amazon Mechanical Turk (table S2) showed that the manipulations produced the intended biasing effects in self-reported similarity (Fig. 2). In particular, changes in responses were most pronounced with regard to the defendants' products, Pieces and OxyClear. In the putatively "Pro-Plaintiff" survey, they received similarity scores that were
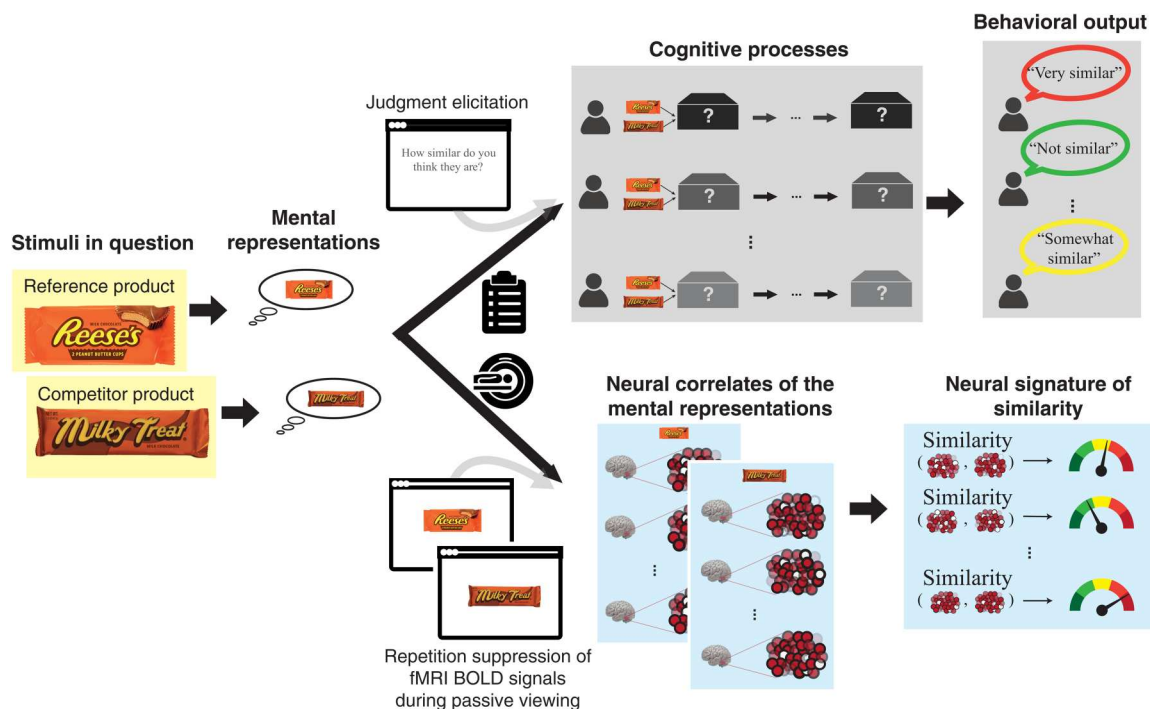
Fig. 1. Improving validation of legal evidence by triangulating mental representations in question. The stimuli in question (e.g., the package designs of two products) generate mental representations, which can be probed by different methods based on distinct assumptions. Survey methods are based on the respondents' own assessments about the relationship between the mental representations and therefore recruit a series of additional cognitive processes. This approach rests on the assumption that these processes are effectively shielded from biases and undue influences. Our neuroscientific approach using RS of the BOLD signals measured by fMRI bypasses these processes, thereby providing a readout of the similarity between stimuli based on the neural correlates of their mental representations. This approach relies on the assumption, among others, of a reliable mapping between such representations and their neural correlates. Voxels in the region of interest (ROI) are represented by dots, and the color shading indicates activation magnitudes.

substantially higher than any other product, painting an exaggerated picture of how much more similar they were to the reference products than other competitors. On the contrary, in the putatively "Pro-Defendant" survey, their reported similarity scores were more or less comparable to those of the other competitor products.

Constructing neural index of visual similarity using RS

To develop a neural index of trademark visual similarity to validate self-report evidence or detect biases therein when contradictory evidence arises, we scanned participants who were blind to the goal of the study using a passive viewing paradigm optimized for RS (Fig. 3A). RS takes advantage of the fact that the neural response declines upon repeated presentation of the same stimulus. Although the underlying neurobiological mechanism remains debated, this phenomenon appears to be a general property of neurons, has been shown to be highly robust across brain regions, and can be observed using different measurement techniques, including fMRI, which measures neural activities indirectly through the BOLD response (24–27). In particular, substantial evidence indicates that the relative suppression between two distinct stimuli can be used to assess the degree of overlap in neural representations of these stimuli (Fig. 3B) (26). Thus, by repeatedly presenting images of different products, we can extract a readout of the mental representation in question, visual similarity, using neural responses from object-sensitive regions of the visual system identified a priori.

In particular, this approach has three important advantages for reducing sources of bias that could be exploited by biased parties in survey-based findings. First, the use of a passive viewing paradigm, i.e., one in which participants do not actively make similarity judgments, allows us to elicit neural responses from participants in the absence of a behavioral response, minimizing "downstream" cognitive processing that can influence subjective report. Second, passive viewing expands our ability to blind subjects to the purpose of the study. Together, these two features permit us to isolate neural responses to the visual stimuli of interest, minimizing the possibility of introducing biases via task instructions and leading questions (e.g., the definition of similarity or which visual features to compare). In contrast, simply administering a survey in the scanner would open the door to the very biases that we are trying to minimize. Third, the use of RS provides us with a readout of the degree of similarity between two stimuli without the need for additional assumptions about how to quantify similarity between representations (35).

To independently identify object-sensitive regions, we first conducted a functional localizer task in which participants were shown object images and scrambled images matched on low-level visual features (Fig. 3C). Specifically, we used the diffeomorphic transformation method developed in (36), which preserves the basic perceptual properties of the image while removing higher-level percepts such as object and category identity and has been shown to be more effective than conventional methods, such as phase
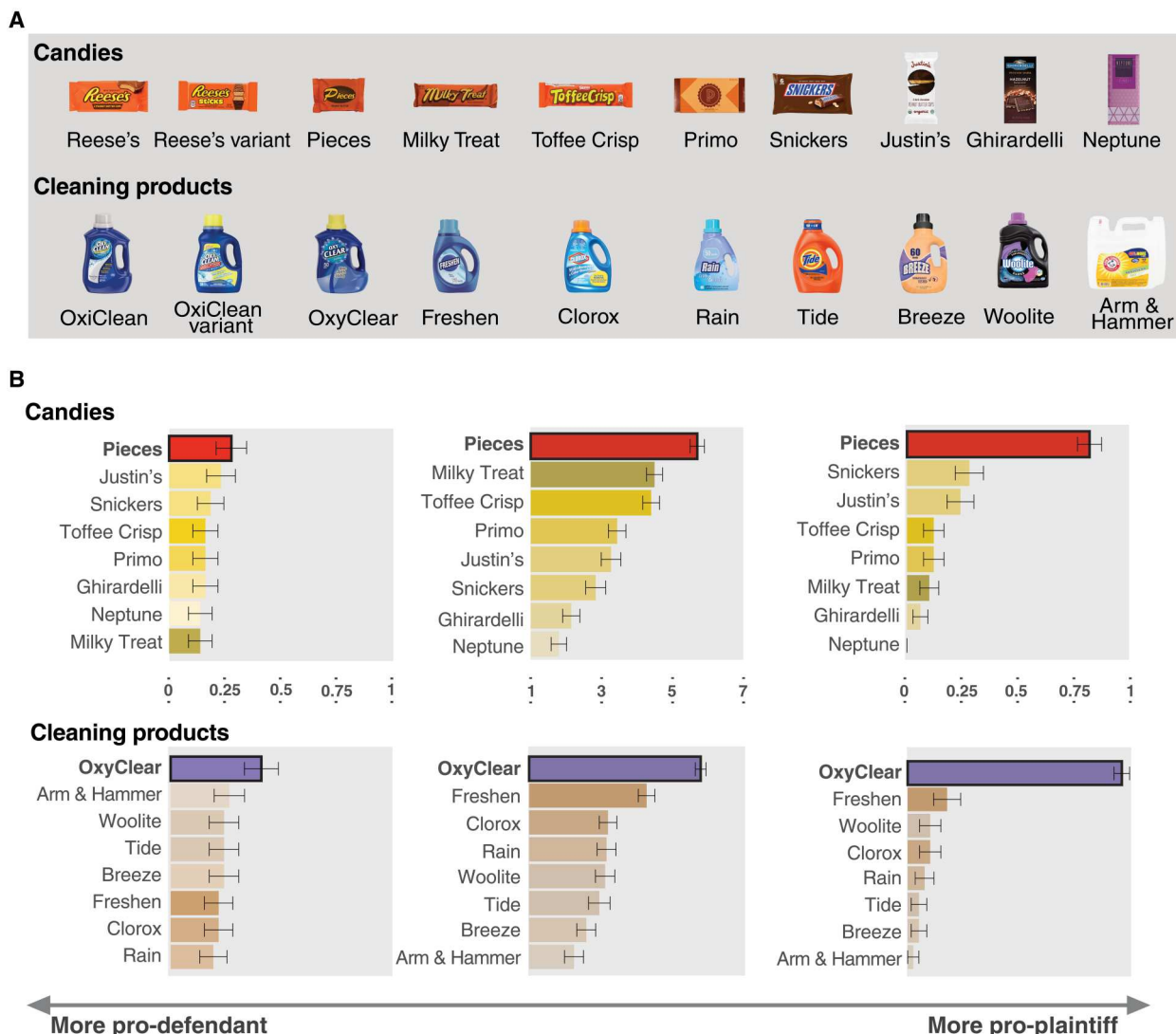
Fig. 2. Contradictory behavioral reports of subjective visual similarity resulting from experimenter-induced biases. (**A**) Full stimulus sets for the candy and cleaning product categories. (**B**) The Pro-Defendant survey (left column) shows significantly lower similarity ratings (x axis) for Pieces and OxyClear, relative to comparison products, than either the Neutral (middle column) or the Pro-Plaintiff (right column) surveys. The latter (right) shows higher subjective similarity ratings for allegedly infringing products Pieces and OxyClear, relative to comparison products, than either the Neutral or the Pro-Defendant surveys. Numbers on the x axes represent either similarity ratings on a 1 to 7 scale (Neutral survey) or the proportion of subjects who judged the competitor product as similar to the reference product (Pro-Defendant and Pro-Plaintiff surveys; see Supplementary Methods).

scrambling and texture scrambling, in ruling out responses to low-level features. Consistent with this approach, a contrast of object versus scrambled images isolated areas in the object-sensitive ventral occipitotemporal cortex, including the fusiform gyrus (table S3), without implicating visual regions involved in the processing of lower-level features.

Next, we extracted the neural responses in the fusiform gyri for the pairs in the main fMRI task and defined a neural similarity index (Fig. 3D) based on a linear transformation and normalization of the raw RS effect (Supplementary Methods), such that values for the index ranged from 0 (greatest activity/weakest RS and, thus, lowest similarity) to 1 (weakest activity/strongest RS and, thus, maximal similarity). For the upper end of the scale, we used the RS effect elicited by consecutive presentations of the same reference

product because the reference product is most similar to itself. For the lower end of the scale, we used the competitor product with the weakest RS effect (see Supplementary Methods). The frequency of occurrence of the reference products and the temporal regularities between the competitor products and the reference products were held constant across all competitor products; thus, these aspects of the task design would not explain any differential RS effects across competitors. Additional analyses also confirmed that different ways of defining the region of interest (ROI) in the fusiform gyri (i.e., individual versus group-based), as well as ROIs of different sizes, yielded consistent similarity indices (fig. S1).
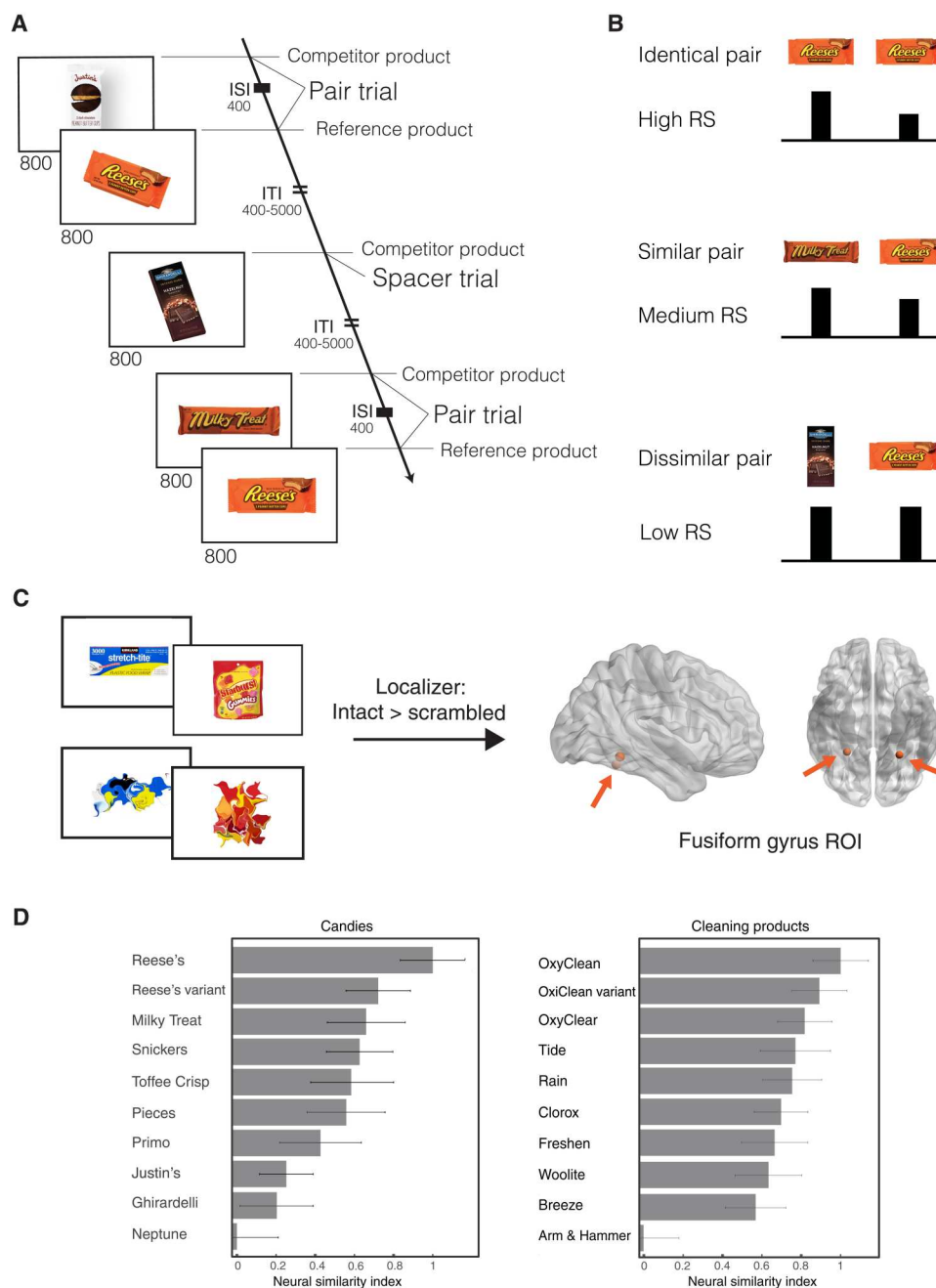
Fig. 3. Measuring perceived similarity with RS. (**A**) fMRI paradigm. Participants viewed a continuous stream of product images that were organized as pair trials (a competitor product followed by the reference product from the same category) and spacer trials (standalone presentations of competitor products). Numbers represent the duration of each phase of the trial in milliseconds, where ISI stands for the interstimulus interval in pair trials and ITI indicates the intertrial interval. During the ISI and ITI, a fixation cross was presented at the center of a white screen (omitted here for clarity). (**B**) Predictions for brain responses in pair trials. For the second (i.e., reference) product, the strength of the neural response, illustrated by the height of the bar, increases as a function of decreasing visual similarity with the first product, i.e., it shows less RS. (**C**) Spherical ROIs for object-sensitive brain areas. These 5-mm-radius ROIs in bilateral fusiform gyrus were defined by a contrast of intact versus scrambled images, as presented during an object localizer task (see Supplementary Methods). (**D**) Neural similarity index for the different products in our stimulus set. The calculation of this neural similarity index based on the fMRI data is described in Supplementary Methods in the section "Brain-based measure of similarity." Error bars indicate SEMs.

Detecting biases in survey instruments using RS

This neural similarity index provides a way to validate similarity judgments elicited by surveys, as well as to detect possible biases introduced into self-report data, intentionally or not, by interested parties (*37*). By examining the consistency between the neural similarity index extracted from the activities of the fusiform gyri ROI (Fig. 4A) and the behavioral measures from each survey (Fig. 4B), we found a strong and statistically significant neural-behavioral correlation for the putatively Neutral survey (Pearson's $r = 0.86$, $P = 0.001$ for candies; $r = 0.71$, $P = 0.02$ for cleaning products). Directionally, consistent with the principle of RS, a higher visual similarity rating between a competitor product and the reference product corresponds to a stronger RS (i.e., higher neural similarity index) when the two products were presented together in the fMRI experiment. In contrast, this high degree of alignment was not observed for either the putatively Pro-Defendant ($r = 0.17$, $P = 0.69$ for

candies; $r = 0.09$, $P = 0.83$ for cleaning products) or the Pro-Plaintiff ($r = 0.40$, $P = 0.33$ for candies; $r = 0.38$, $P = 0.35$ for cleaning products) survey (Fig. 4B), indicating that the behavioral responses in these surveys were not well supported by the neural index.

To formally evaluate the relative alignment of pairs of surveys against the neural similarity index (Supplementary Methods), we measured the mean square distance (MSD) between the neural similarity index and the normalized behavioral similarity score for each survey. In both categories, the MSD of the putatively neutral survey was significantly lower than those of the putatively Pro-Plaintiff and Pro-Defendant surveys ($P < 0.001$ for both Neutral versus Pro-Plaintiff and Neutral versus Pro-Defendant in both categories; Fig. 4C), suggesting that the neural similarity index is indeed capable of distinguishing between surveys containing different amounts of bias.
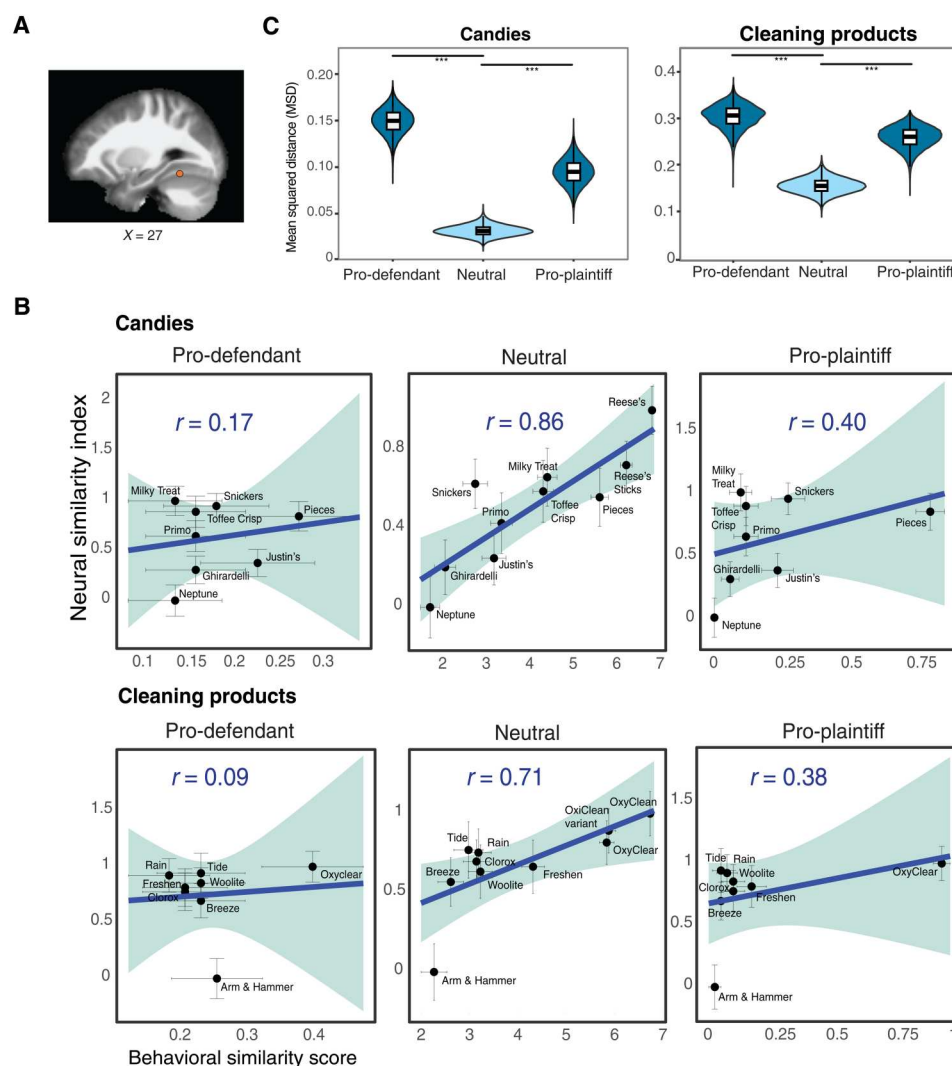

Fig. 4. Using the neural similarity index as a benchmark to compare contradictory behavioral reports. (**A**) The fusiform ROI shown on a sagittal plane ($X = 27$; only the right side is shown). (**B**) Scatter plots of the neural similarity index versus the behavioral similarity score from each version of the surveys. Best-fit linear regression lines are shown in blue, and the Pearson's r is included. Shaded areas indicate 95% confidence bands. (**C**) Violin plots of the distribution of brain-behavior distance for each survey. Distance was calculated as the MSD between normalized neural and behavioral measures using a bootstrap resampling procedure. The box plot within each violin further displays the median and interquartile range. ***$P < 0.001$.

Because the neural index was derived from brain activities using a behaviorally orthogonal (passive viewing) task in which no similarity judgments were elicited from the participants, the alignment between the neural similarity index and the ratings in the putatively Neutral survey cannot be explained as the result of a simple mapping between the explicit similarity judgments and the brain activities while such judgments were made. In addition, in a whole-brain analysis, responses to similarity were concentrated in the bilateral fusiform gyrus (Supplementary Methods, fig. S2, and table S4), lending further support to our focus on activities in the object processing regions of the brain to derive the neural similarity index.

Replicability and sensitivity of RS index

Last, we sought to test the replicability and sensitivity of our findings by asking whether our neural index is capable of detecting more subtle forms of bias. Specifically, we removed some of the more prominent manipulations in our surveys, such as references to copycats and "bullies" (table S1 and Materials and Methods). As expected, the intended biasing effects were diminished in the putatively Pro-Plaintiff and Pro-Defendant surveys that were modified (Fig. 5A). In the former, while the defendant's products still received the highest similarity scores, the difference with the other competitor products was less marked, particularly in the candy category. Likewise, the similarity scores for the defendant's products stood out to a larger extent in the Pro-Defendant surveys. At the same time, the general biasing effects of the Pro-Plaintiff and Pro-Defendant surveys persisted, and the overall pattern of the behavioral responses to the different surveys remained similar to the previous results.

Using the same procedure as the previous study, we found that in both categories, the MSD of the neutral survey remained significantly lower than those of the Pro-Plaintiff and Pro-Defendant surveys ($P < 0.001$ for Neutral versus Pro-Defendant and $P = 0.02$ for Neutral versus Pro-Plaintiff in candies; $P < 0.001$ for both comparisons in cleaning products; Fig. 5B). Thus, even in the case of the Pro-Plaintiff survey for candies, where the ratings are qualitatively similar to those in the Neutral survey, our neural index was able to identify the latter as the more accurate survey, albeit at a lower level of confidence. Nevertheless, these results provide support for the idea that our neural similarity index can identify even fairly nuanced forms of bias in self-report evidence.

DISCUSSION

Legal tests invoking the viewpoint of average, ordinary, or typical persons play an important role in nearly every area of the legal system (3–10). Despite vast differences in the application of such tests, legal scholars underscore a shared need for more robust ways of applying them (3–10). In particular, concerns about an overreliance on either intuition or evidence of uncertain reliability (4, 6, 10) have led to criticism in some quarters that such tests exist primarily "in the eye of the beholder" (3, 5). Even those more favorably disposed to reasonableness standards recognize that their inherent flexibility comes at the cost of possibly inconsistent outcomes and greater potential for bias in the decision-making process (38).

Here, we sought to address the need articulated by judges and legal scholars to better align this class of legal tests with their goals (11, 39). Specifically, using aggregate neural signatures of relevant mental states as an empirically grounded measure, we address, at least within the parameters of the present experiment, some common conundrums facing the court and the litigants: How should one assess the strength of putative evidence about the mental state of some idealized group of individuals? When the pieces of evidence presented by the opposing parties are contradictory, how should one determine their respective credibility?

A key strength of our approach to these questions lies in the use of RS to give a readout of the representational similarity between stimuli (26) that avoids potential biases introduced, either intentionally or unintentionally, during self-report elicitation, e.g., via use of leading questions. This technique allows us to elicit participant responses even when they are blind to the purpose of the study and even without explicit instructions on the nature of the desired comparisons (11, 40). Consequently, these features allow us to directly address concerns over (i) the ability of respondents to prevent normative aspects of the question to color their judgments and (ii) the ability of motivated litigants to exploit these vulnerabilities, as we did in our biased instruments.

Theoretical and practical considerations

Such results represent a modest but concrete step toward improving the relevance of neuroscientific findings for the law. On the one hand, our approach addresses only a specific, albeit common, instance of the broader legal need to determine the aggregate response of a legally relevant population. We do not claim or anticipate that this approach can provide a comprehensive test generalizable to other legal questions. On the other hand, particularly given the scarcity of existing empirical work in this area, we emphasize the potential significance of a more scientific measure of legally relevant mental states and the benefits of considering legal questions that go beyond G2i inferences (17, 18).

In particular, we draw on two challenges identified in a pioneering study by Vilares and colleagues (41), who sought to identify "culpable mental states" in criminal law ranging from purpose and knowledge to recklessness and negligence. First, they note that the relevant mental states may no longer exist at the time of the study or might exist in a much-altered form, for example, when they represent the memory of a mental state rather than the state itself. In such cases, neural data may need to capture the person's previous mental state over a time frame ranging from the recent to the distant past, a task that can be exceedingly challenging. This concern is reduced in our study, in which the alleged infringement reflects ongoing perceptions as opposed to a previous, one-time event. Moreover, as also true in our case, it is unlikely that visual processing is particularly malleable, but this assumption may not hold in other instances.

The second challenge concerns the issue of the representativeness of our sample. This issue is unexpectedly complex, as the population of inference can vary greatly. For example, in trademark law alone, the relevant population may consist of all consumers, all consumers of a particular mark, or consumers within a particular market if the mark in question is distributed only locally. Therefore, although we are able to assess the internal consistency of our data (fig. S3), we do not provide guidance on what should be the population of inference, which depends on normative concerns orthogonal to evidence provided by our study.
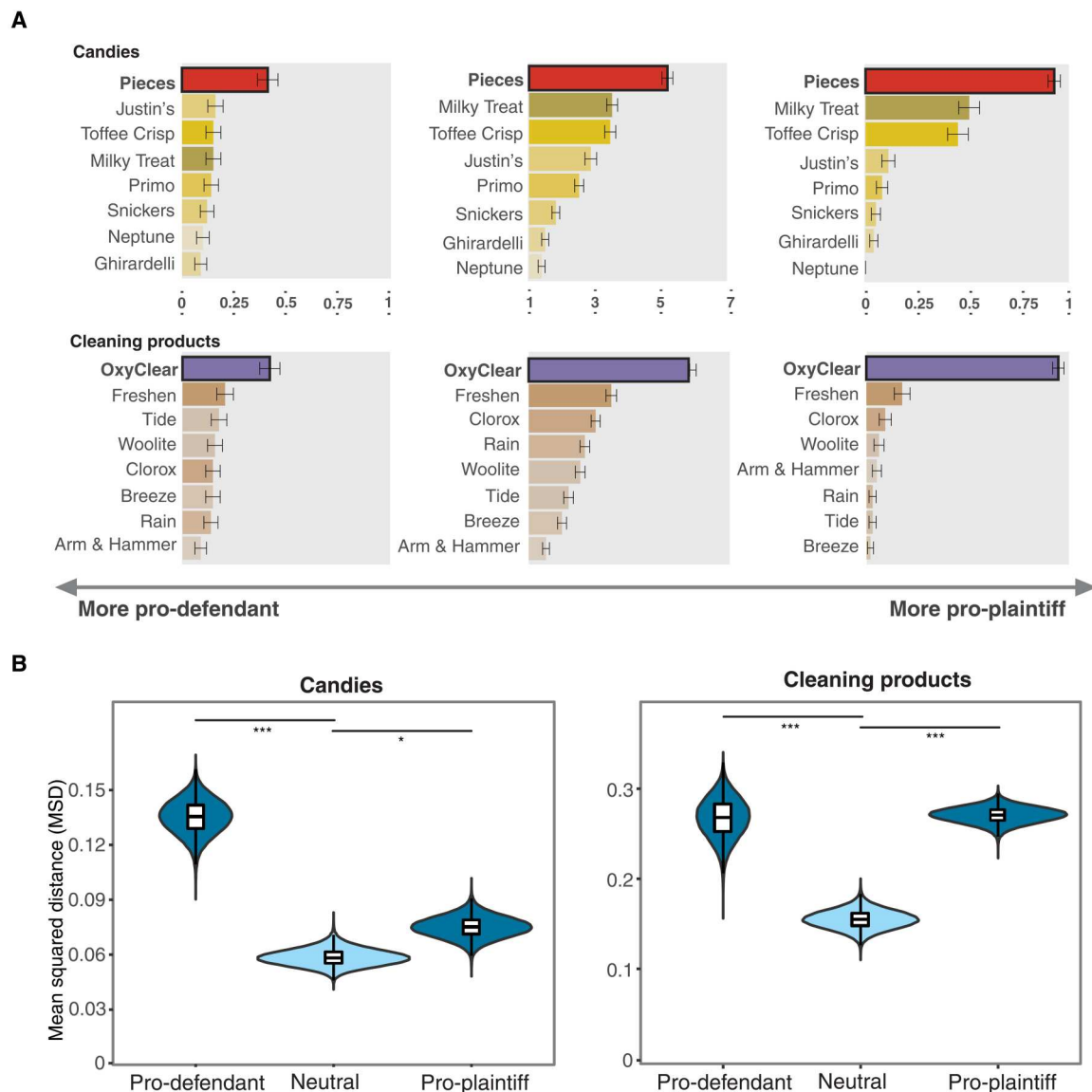
Fig. 5. Assessing robustness and sensitivity of neural similarity index as benchmark. (**A**) Behavioral reports that replicated the direction of bias in Fig. 3 but with diminished magnitude. Format identical to Fig. 2. (**B**) Violin plots of the distribution of brain-behavior distance for each survey. Format is identical to Fig. 4B. *$P < 0.05$; ***$P < 0.001$.

In a related fashion, it is important to note that our results are not capable of providing normative guidance about when one mark is "too similar" to another. Rather, as is typical with scientific evidence used by the court, our methodological approach should be seen as presenting a basis for a practical and workable test of determining similarity (28, 42) while remaining silent about the existence of infringement (or the lack thereof). Ultimately, the similarity threshold that determines the existence of infringement is not an empirical question but a normative one and hence requires input from legal theorists and practitioners.

Moreover, as in other types of technical and scientific evidence involving a significant degree of expertise, it is important for judges to serve as gatekeepers to determine what is admissible, to instruct jurors on how to determine the appropriate weight to place on the evidence, and to ensure that the probativity of different forms of evidence is not outweighed by their potential prejudicial effects (43–46). This role is particularly important in light of findings suggesting that jurors may overweigh the evidential value of brain-based evidence, in part because of the visual and intuitive appeal of brain images (47), although other studies have challenged these findings (48).

Open questions and future directions
Beyond the set of general challenges discussed above, there also exist issues specific to our application case that can benefit from additional investigation. Although visual similarity is the primary driver of consumer confusion in many cases (49), it is only one of a set of criteria used by the court to determine infringement. Similar to litigation surveys, our method relies on an assumption that the specific display chosen is representative of the product's appearance

in the real world. It is possible that similarity and confusion may vary across situations, for example, where certain features appear more salient in certain contexts. How to define ecologically valid circumstances and how to determine similarity and confusion in ecologically valid circumstances remain important questions for future investigation. Future studies could further broaden the value of neuroscientific evidence by investigating factors including conceptual similarity (50), individual differences in proneness to confusion (51), and time constraints (52), which are also known to contribute to consumer confusion (52, 53).

In addition to consumer surveys, a number of approaches have been proposed that seek to provide a holistic measure of consumer confusion, including those using showcards (53), in-store interviews (54), slides (49, 52), the tachistoscope (52), a coupon redemption test (55), and image blurring (56). Although a full review of this literature is beyond the scope of the current paper, a common strength of these methods is their potential to provide a more direct measure of confusion, which is not possible using consumer surveys or our proposed method. Despite this advantage, holistic measures of confusion have thus far seen limited use, either in absolute terms or in comparison to consumer surveys. This state of affairs is likely for two reasons. First, the sheer diversity of the proposed measures suggests considerable challenges facing standardization efforts, even more so than those involving consumer surveys. Second, direct measures of consumer confusion provide limited insight into the sources of confusion, a critical component for satisfying the legal criteria determining infringement. Nevertheless, given their distinct strengths and weaknesses, it is entirely possible that a mixed approach combining holistic measures with consumer surveys and neural data would be stronger than each separately.

Substantial future work is also needed to determine appropriate legal standards for the collection, analysis, and interpretation of neuroimaging evidence going forward. These standards would include characterizing the specific task paradigm that best addresses the mental representation at issue and identifying the number of subjects required to achieve acceptable test-retest reliability. Studies with a larger set of products could also help better quantify the precise relationship between neural and self-reported measures of similarity, as well as the possible presence of nonlinearities. Inclusion of longer interstimulus intervals could allow for data-driven estimation of the hemodynamic response function (HRF), as opposed to the standard canonical HRF used in the present study (57). Such standardization would be an important step toward enabling detection of less transparent forms of bias and improving generalizability of our methods to real-world cases. Biases in real legal cases may be more subtle than those studied here. At the same time, as demonstrated by the reproducibility literature, substantial bias can result from a combination of individual analytical choices that appear defensible in isolation (34). Thus, at a minimum, our method provides a first step toward using neural data to limit the degree of bias to which self-report evidence is vulnerable.

Given specific legal standards, it will also be important to verify the validity of such inferences. Doing so can be a considerable challenge given that inner workings of the respective parties that produced findings suspected of bias are not known. That is, it may not be possible to verify post hoc that our assessments of relative bias are accurate. One possible solution, borrowed from the statistical fraud detection literature, is to implement an auditing and retesting procedure for instruments suspected of bias (58, 59). This type of procedure, analogous to a replication study in scientific research, has found success in identifying cheating on standardized testing by showing that classrooms suspected of cheating experienced large declines in test scores when retested under controlled conditions (59).

More generally, our work can contribute by shedding light on opportunities and challenges of applying neuroscientific data to other instances in which the law must assess the reactions of a particular demographic category. In copyright law, for example, a key legal test is the extent to which two works are "substantially similar" from the vantage point of the "ordinary observer" (6, 40). Beyond intellectual property law, obscenity law is another legal domain in which survey evidence has been used to assess whether the public thought a publication was obscene (60). A similar perception exists with respect to the capriciousness and susceptibility to bias of such assessments. Copyright scholars, for example, describe the ordinary observer test's application as "artificial and disappointingly inaccurate" (40). Tort law asks judges and juries to determine whether a defendant's behavior falls below the standard of a reasonable person, and critics of obscenity law argue that "judges and juries are left to create their own standard in each case" (61).

Methodological considerations
Given these similarities, future neuroscientific studies could help address these and other articulated needs by developing measures that can serve as the basis of practical and workable tests. At the same time, such studies must account for a host of challenges not present in our study, and it may well be that methods other than RS will be more appropriate in these circumstances. A number of candidate neural recording modalities exist to probe and characterize neural representations, such as fMRI, electroencephalography, and magnetoencephalography, each of which captures different elements of the neural response; and different analytic tools, such as univariate methods, multivoxel pattern analysis (MVPA), and representational similarity analysis (RSA), may be useful depending on how localized or distributed the neural representation might be.

In the present study, our choice of RS was motivated by its long track record in visual neuroscience of quantifying the similarity of visual representations, particularly those representations found in primary, secondary, and association visual cortices (26, 27, 62). This wealth of prior information provided an important benefit in reducing the number of design and analytical choices that we needed to make, all of which could be considered to be "experimenter degrees of freedom" from a legal perspective. From a technical perspective, the relative simplicity of RS, which allows experimenters to circumvent the difficult problems of accounting for individual differences in the spatial pattern of neural activities (63) and specifying the functional form of (dis)similarity, may also offer certain advantages in reducing bias in a legal setting over other analytic tools such as MVPA and RSA.

On the other hand, some of the methodological requirements for RS may limit the generality of its application. Whereas RS requires rapid presentation of stimuli and focuses on the relationship between stimuli, the ability of RSA and MVPA to focus on individual stimuli provides them with an analytic flexibility for subsequent comparisons that makes them attractive to a number of research questions. The latter methods may also have an advantage in situations involving more abstract and distributed representations, such

as concepts (64), as RS, in comparison, is best used in cases where there exists a clearly defined or well-circumscribed neuroanatomical substrate. This advantage may be important in a wide range of legal applications, given that typically complex real-world stimuli likely recruit representations across multiple brain regions. More broadly, RS, MVPA, and RSA may differentially capture aspects of representational similarity arising from different neural mechanisms. Understanding how they map onto substantive behavioral phenomena, such as confusability (65), remains an active area of research, the growing insights from which will certainly better inform the choice of analytical approach for broader applications. Understanding the relative strengths and weaknesses of different neuroscientific methods, as well as the degree to which they are complementary, in addressing needs of different legal applications will be an exciting next step in the application of neuroscientific methods to the law.

Stronger together than apart

Last, given our findings, one may well be tempted to ask why the courts should not simply request a well-crafted, neutral survey. Although possible in principle, such requests necessarily run into the very real problem that there does not exist a generic set of standards specifying what constitutes a "neutral" or "well-crafted" survey. Consequently, it can be exceedingly difficult for the courts to separate serious claims about bias from frivolous ones, particularly under an adversarial system. Moreover, this difficulty has contributed to a certain degree of cynicism regarding the value of evidence and expert witnesses as a whole. No less than the eminent legal scholar R. Posner remarked in a legal opinion that, "Many experts are willing for a generous (and sometimes for a modest) fee to bend their science in the direction from which their fee is coming" (66).

Conversely, one could equally ask why there is a need for self-report data at all. Here, too, we note important advantages that self-report measures retain over neural measures. For example, one can typically obtain larger samples due to the lower costs involved. Consumer surveys are also more flexible in that they can be applied to cases where the neural representations are poorly understood or difficult to access, e.g., the very notion of "consumer confusion" itself (67). Thus, perhaps the most important contribution of neuroscientific data lies in its ability to provide an independent benchmark that can limit bias, either directly or indirectly by enabling auditing and retesting procedures (59).

This possibility is consistent with what Morse (42) refers to as an "internalizing" strategy in which the law adopts scientific criteria as legal criteria, as opposed to an "externalizing" strategy in which scientific experts are asked to pass normative judgments. It also accords well with the framework developed by Jones (28) that describes ways in which neuroscience may aid the law, particularly via detecting (mental states in question), buttressing (other forms of evidence), and challenging (evidence of questionable quality). Although still highly imperfect and incomplete, even the small step we take here may constitute a productive and meaningful advance, given the ubiquity of and the acknowledged flaws in current efforts to apply those standards.

MATERIALS AND METHODS

Data and main analysis scripts are available at the following Open Science Framework repository: https://osf.io/n328c/?view_

only=40204543199b4fb69503781265788cca. The fMRI contrast maps are also available through the following NeuroVault repository: https://neurovault.org/collections/NVNLEYZN/. Full methodological details are provided in Supplementary Methods.

Participants

A total of 26 (16 females; mean age, 22.0 ± 4.9 years; range, 18 to 39 years; all right-handed) and 870 individuals participated in the fMRI and the behavioral studies, respectively (details are provided in Supplementary Methods and table S2). This research was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley. Participants provided written informed consent before participation.

Introducing and detecting bias in survey responses

To create experimenter-induced biases in behavioral measures of visual similarity, we designed three versions of surveys (table S1): a putatively Neutral survey, a putatively Pro-Defendant survey that aimed to create responses that favor the hypothetical, highly similar defendant products (Pieces and OxyClear), and a putatively Pro-Plaintiff survey that aimed to favor the proposed plaintiffs (Reese's and OxiClean). Detailed descriptions and a link to the full surveys can be found in Supplementary Methods. Two samples were collected with these surveys, with the second sample using modified versions of the putatively Pro-Defendant and Pro-Plaintiff surveys that removed some of the more prominent manipulations (table S1). To use the neural similarity index as a benchmark to assess the relative bias in the different surveys, the mean squared distance between each survey and the neural measure was computed, and a bootstrap procedure with 100,000 samples was used to determine statistical significance (Supplementary Methods).

Procedure

fMRI participants undergoing neuroimaging completed eight scanning runs of the main task paradigm (Fig. 1A), alternating between two product categories, and one run of the object localizer task (Fig. 1C; Supplementary Methods). In total, each category consisted of 10 products, including the reference product itself, a product variant from the same manufacturer, and four fictitious products plus four other real brands of varying similarity to the reference product.

During the task, participants viewed rapid presentations of product images, shown randomly in one of three possible viewing angles (Fig. 3A). Each of the competitor products and the reference product were grouped to form category-specific pairs in which a competitor product was followed by the reference product after a short interval. Pairs with two consecutive presentations of the reference product were also included. Additional single presentations of the competitor products ("spacer trials") were randomly interleaved to minimize appearance of temporal regularities in the presentation of the pairs.

Participants were not aware of the background or the purpose of the study. Instead, they performed an unrelated task in which they pressed a button every time they saw an inverted image, which appeared on average every nine trials in a pseudorandom manner (Supplementary Methods).

Behavioral participants were recruited from Amazon Mechanical Turk in two separate waves for three different versions of

similarity surveys that were intended to induce biases in different directions in the responses (Supplementary Methods). No participant underwent more than one survey version.

fMRI data analysis

Details of the fMRI acquisition and analysis are provided in Supplementary Methods. Event-related analyses of fMRI time series were performed in the ROI identified from the object localizer task. Regressors were convolved with the canonical HRF. In the main analysis, a neural similarity index was defined on the basis of fMRI RS using BOLD signals for each product. Specifically, we evaluated the corresponding neural activations during the main task when each product was paired with the reference product in the category.

Supplementary Materials

This PDF file includes:
Supplementary Methods
Figs. S1 to S3
Tables S1 to S4
References

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. *Williams* v. *Gaye*, 895 F.3d 1106 (9th Cir., 2018).
2. *Colgate-Palmolive Company* v. *J.M.D. All-Star Import*, 486 F. Supp. 2d 286 (Dist. Court, 2007).
3. L. Heymann, The reasonable person in trademark law. *St. Louis Univ. Law J.* **52**, 781–794 (2008).
4. J. Gardner, The many faces of the reasonable person. *Law Q. Rev.* **131**, 563–584 (2016).
5. B. C. Zipursky, Reasonableness in and out of negligence law. *Univ. Pa. Law Rev.* **163**, 2131–2170 (2015).
6. I. D. Manta, Reasonable copyright. *Boston Coll. Law Rev.* **53**, 1303–1355 (2012).
7. M. Moran, The reasonable person: A conceptual biography in comparative perspective. *Lewis Clark Law Rev.* **14**, 1233–1284 (2010).
8. E. Jaeger, Obscenity and the reasonable person: Will he know it when he sees it. *Boston Coll. Law Rev.* **30**, 823–839 (1988).
9. E. D. DeRosia, Fixing ever-ready: Repairing and standardizing the traditional survey measure of consumer confusion. *Georgia Law Rev.* **53**, 613–682 (2019).
10. B. Beebe, An empirical study of the multifactor tests for trademark infringement. *Calif. Law Rev.* **94**, 1581–1654 (2006).
11. M. Bartholomew, Neuromarks. *Minn. Law Rev.* **103**, 521–585 (2018).
12. S. S. Diamond, D. J. Franklyn, Trademark surveys: An undulating path. *Tex. Law Rev.* **92**, 2029–2047 (2013).
13. F. van Horen, R. Pieters, When high-similarity copycats lose and moderate-similarity copycats gain: The impact of comparative evaluation. *J. Market. Res.* **49**, 83–91 (2012).
14. J. L. Zaichkowsky, *The Psychology Behind Trademark Infringement and Counterfeiting* (Psychology Press, 2006).
15. *Polaroid Corporation* v. *Polarad Electronics Corp.*, 287 F.2d 492 (Court of Appeals, 2nd Circuit, 1961).
16. B. Beebe, R. Germano, C. J. Sprigman, J. H. Steckel, Testing for trademark dilution in court and the lab. *Univ. Chic. Law Rev.* **86**, 611–668 (2019).
17. J. W. Buckholtz, D. L. Faigman, Promises, promises for neuroscience and law. *Curr. Biol.* **24**, R861–R867 (2014).
18. D. L. Faigman, J. Monahan, C. Slobogin, Group to individual (G2i) inference in scientific expert testimony. *Univ. Chic. Law Rev.* **81**, 417–480 (2014).
19. D. Purves, et al., *Principles of Cognitive Neuroscience*, (Sinauer Associates, ed. 2, 2012).
20. D. Pins, M.-E. Meyer, J. Foucher, G. Humphreys, M. Boucart, Neural correlates of implicit object identification. *Neuropsychologia* **42**, 1247–1259 (2004).
21. M. Valdés-Sosa, M. Ontivero-Ortega, J. Iglesias-Fuster, A. Lage-Castellanos, J. Gong, C. Luo, A. M. Castro-Laguardia, M. A. Bobes, D. Marinazzo, D. Yao, Objects seen as scenes: Neural circuitry for attending whole or parts. *Neuroimage* **210**, 116526 (2020).
22. C. E. Wierenga, W. M. Perlstein, M. Benjamin, C. M. Leonard, L. G. Rothi, T. Conway, M. A. Cato, K. Gopinath, R. Briggs, B. Crosson, Neural substrates of object identification:

23. J. Zhang, X. Li, Y. Song, J. Liu, The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLOS ONE* **7**, e40390 (2012).
24. X. Jiang, E. Rosen, T. Zeffiro, J. VanMeter, V. Blanz, M. Riesenhuber, Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron* **50**, 159–172 (2006).
25. P. Rotshtein, R. N. A. Henson, A. Treves, J. Driver, R. J. Dolan, Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* **8**, 107–113 (2005).
26. H. C. Barron, M. M. Garvert, T. E. J. Behrens, Repetition suppression: A means to index neural representations using BOLD? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150314–20150355 (2016).
27. K. Grill-Spector, R. Henson, A. Martin, Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23 (2006).
28. O. D. Jones, Seven ways neuroscience aids law, in *Neurosciences and the Human Person: New Perspectives on Human Activities*, A. Battro, S. Dehaene, W. Singer, Eds. (Scripta Varia, 2013).
29. S. Balganesh, I. D. Manta, T. Wilkinson-Ryan, Judging similarity. *Iowa Law Rev.* **100**, 267–296 (2014).
30. *Hershey Company* v. *Posh Nosh Imports (USA) Inc.*, No.2: 14-cv-04028 (Dist. Court California, 2014).
31. National Research Council, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009).
32. I. Simonson, Trademark infringement from the buyer perspective: Conceptual analysis and measurement implications. *J. Public Policy Mark.* **13**, 181–199 (1994).
33. I. Simonson, R. Kivetz, Demand effects in likelihood of confusion surveys: The importance of conditions marketplace, in *Trademark and Deceptive Advertising Surveys* (ABA Book Publishing, 2012), pp. 1–18.
34. U. Simonsohn, L. D. Nelson, J. P. Simmons, P-curve: A key to the file-drawer. *J. Exp. Psychol. Gen.* **143**, 514–534 (2014).
35. A. Walther, H. Nili, N. Ejaz, A. Alink, N. Kriegeskorte, J. Diedrichsen, Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).
36. B. Stojanoski, R. Cusack, Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *J. Vis.* **14**, 6 (2014).
37. R. Bird, J. Steckel, The role of consumer surveys in trademark infringement: Empirical evidence from the federal courts. *U. Pa J. Bus. Law* **14**, 1013–1055 (2017).
38. K. M. Sullivan, Foreword: The justices of rules and standards. *Harv. Law Rev.* **106**, 22–123 (1992).
39. R. Tushnet, Gone in sixty milliseconds: Trademark law and cognitive Science. *Tex. Law Rev.* **86**, 507–569 (2008).
40. M. Bartholomew, Copyright and the brain. *Wash. Univ. Law Rev.* **98**, 525–587 (2020).
41. I. Vilares, M. J. Wesley, W. Y. Ahn, R. J. Bonnie, M. Hoffman, O. D. Jones, S. J. Morse, G. Yaffe, T. Lohrenz, P. R. Montague, Predicting the knowledge-recklessness distinction in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3222–3227 (2017).
42. S. J. Morse, The status of neurolaw: A plea for current modesty and future cautious optimism. *J. Psychiatry Law* **39**, 595–626 (2011).
43. V. P. Hans, Judges, juries, and scientific evidence. *J. Law Policy* **16**, 3 (2008).
44. T. Daftary-Kapur, R. Dumas, S. D. Penrod, Jury decision-making biases and methods to counter them. *Legal Criminol. Psychol.* **15**, 133–154 (2010).
45. E. J. Imwinkelried, Judge versus jury: Who should decide questions of preliminary facts conditioning the admissibility of scientific evidence. *William Mary Law Rev.* **25**, 577 (1983).
46. D. Aono, G. Yaffe, H. Kober, Neuroscientific evidence in the courtroom: A review. *Cogn. Res. Princ. Implic.* **4**, 40 (2019).
47. D. P. McCabe, A. D. Castel, Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition* **107**, 343–352 (2008).
48. C. J. Hook, M. J. Farah, Look again: Effects of brain images and mind–brain dualism on lay evaluations of research. *J. Cogn. Neurosci.* **25**, 1397–1405 (2013).
49. B. Loken, I. Ross, R. L. Hinkle, Consumer "confusion" of origin and brand similarity perceptions. *J. Public Policy Mark.* **5**, 195–211 (1986).
50. F. van Horen, R. Pieters, Consumer evaluation of copycat brands: The effect of imitation type. *Int. J. Res. Mark.* **29**, 246–255 (2012).
51. G. Walsh, T. Hennig-Thurau, V.-W. Mitchell, Consumer confusion proneness: Scale development, validation, and application. *J. Mark. Manag.* **23**, 697–721 (2007).
52. J. Kapferer, Brand confusion: Empirical study of a legal concept. *Psychol. Mark.* **12**, 551–568 (1995).
53. G. Balabanis, S. Craven, Consumer confusion from own brand lookalikes: An exploratory investigation. *J. Mark. Manag.* **13**, 299–313 (1997).

54. G. Miaoulis, N. d'Amato, Consumer confusion & trademark infringement: Presents a new, broadened concept of consumer confusion, illustrated by research results in the Tic Tac case. *J. Mark.* **42**, 48–55 (1978).

55. R. B. Boal, Techniques for ascertaining likelihood of confusion and the meaning of advertising communications. *Trademark Report.* **73**, 405–435 (1983).

56. J.-N. Kapferer, Stealing brand equity: Measuring perceptual confusion between national brands and "copycat" own-label products. *Mark. Res. Today* **23**, 96–102 (1995).

57. K. J. Friston, C. D. Frith, R. Turner, R. S. J. Frackowiak, Characterizing evoked hemodynamics with fMRI. *Neuroimage* **2**, 157–165 (1995).

58. R. J. Bolton, D. J. Hand, Statistical fraud detection: A review. *Stat. Sci.* **17**, 235–255 (2002).

59. B. Jacob, S. Levitt, To catch a cheater. *Educ. Next* , 69–75 (2003).

60. J. T. Clark, The "Community Standard" in the trial of obscenity cases—A mandate for empirical evidence in search of the truth. *Ohio N.U.L. Rev.* **20**, 13–56 (1993).

61. B. L. Newman, Constitutional Law—The problem with obscenity. *Case West Res. Law Rev.* **11**, 612–669 (1959).

62. R. Sayres, K. Grill-Spector, Object-selective cortex exhibits performance-independent repetition suppression. *J. Neurophysiol.* **95**, 995–1007 (2006).

63. F. M. Ramírez, C. Revsine, E. P. Merriam, What do across-subject analyses really tell us about neural coding? *Neuropsychologia* **143**, 107489 (2020).

64. C. B. Martin, D. Douglas, R. N. Newsome, L. L. Man, M. D. Barense, Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife* **7**, e31873 (2018).

65. M. Hatfield, M. McCloskey, S. Park, Neural representation of object orientation: A dissociation between MVPA and repetition suppression. *Neuroimage* **139**, 136–148 (2016).

66. *Indianapolis Colts* v. *Metro. Balt. Football Club*, 34 F. 3d 410, 416 (7th Cir., 1994).

67. R. Bryan, La Victoria Foods v. Curtis-Burns, in *Social Science in Law: Cases and Materials*, J. Monahan, L. Walker, Eds. (Foundation Press, ed. 3, 1994).

68. R. H. Thornburg, Trademark survey evidence: Review of current trends in the Ninth Circuit. *Santa Clara High Technol. Law J.* 21, 715 (2004).

69. N.-C. Woods, Survey evidence in Lanham Act violations. *Trinity Law Rev.* **15**, 67 (2008).

70. C. Summerfield, E. H. Trittschuh, J. M. Monti, M.-M. Mesulam, T. Egner, Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11**, 1004–1006 (2008).

71. A. C. Jenkins, C. N. Macrae, J. P. Mitchell, Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4507–4512 (2008).

72. J. D. Power, B. M. Silver, M. R. Silverman, E. L. Ajodan, D. J. Bos, R. M. Jones, Customized head molds reduce motion during resting state fMRI scans. *Neuroimage* **189**, 141–149 (2019).

73. M. Brett, J.-L. Anton, R. Valabregue, J.-B. Poline, Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage* **16**, S497 (2002).