THIS IS A PREPRINT

Neal, T.M.S., MacLean, N., Morgan, R.D., & Murrie, D.C. (2024). Confirmatory information seeking is robust in psychologists' diagnostic reasoning. *Law and Human Behavior, 48*(5-6), 503-518. https://doi.org/10.1037/lhb0000574

Preprint: https://osf.io/preprints/psyarxiv/6rhck

Confirmatory Information Seeking Is Robust in Psychologists' Diagnostic Reasoning

Tess M. S. Neal[1], Nina MacLean[2], Robert D. Morgan[3], and Daniel C. Murrie[4]

[1] Department of Psychology, Iowa State University

[2] Michigan Department of Health & Human Services, Saline, Michigan, United States

[3] College of Health and Human Sciences, Southern Illinois University

[4] Institute of Law, Psychiatry, and Public Policy, University of Virginia

Author Note

Tess M. S. Neal https://orcid.org/0000-0002-9528-8638

Nina MacLean https://orcid.org/0000-0002-2689-9240

Robert D. Morgan https://orcid.org/0000-0002-6263-5955

Daniel C. Murrie https://orcid.org/0000-0002-4195-1238

At the time these data were collected, Tess M. S. Neal was affiliated with Arizona State University and Nina MacLean and Robert D. Morgan were affiliated with Texas Tech University.

The materials, data, and code are available on the Open Science Framework (https://osf.io/678cs/).

We presented portions of these data at the 2019 annual conference of the Society for Personality and Social Psychology in Portland, Oregon; the 2018 annual conference of the Law and Society Association in Toronto, Ontario, Canada; and the 2017 and 2018 annual conferences

of the American Psychology-Law Society in Seattle, Washington, and Memphis, Tennessee, respectively. The two-paragraph description of the development of our participant database in Study 1 of this article first appeared in the work by MacLean et al. (2019) and is reproduced here per the American Psychological Association's Fair Use Criteria. MacLean et al. reported results from other survey data collected at the same time as experimental data from Study 1 reported in the current article. All experimental manipulations for Study 1 appear only in this article. During the collection of data for Study 1, the experimental portion of the project appeared before the survey materials, and more people responded to the experimental portion of that data collection effort than the survey questions. Thus, we report a higher number of participants in the experimental data here in Study 1 ($N = 149$) than we did for the survey portion of that data collection effort as reported by MacLean et al. (2019; $N = 120$). We have no known conflicts of interest to disclose.

The American Psychology-Law Society and the Society for the Psychological Study of Social Issues supported this research. The National Science Foundation supported the first author's work on this project under Grant 1655011. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or other funders. We thank Michael J. Saks for his feedback on the design of Studies 1 and 2; Hannah Goddard, Jeffrey Haas, Cassandra Hamza, Morgan Hight, Brian Howatt, Joseph Hunter, Sierra Marshall, Brieann Olafsson, Carina Philipp, Caitlyn Quamme, Dennis Ramos, Bibiana Oliveira, Cassidy Siso, and Rebecca Velez for their assistance in creating the participant database; and Philip J. Neal for his help in preparing the manuscript for publication.

Correspondence concerning this article should be addressed to Tess M. S. Neal, Department of Psychology, Iowa State University, 1347 Lagomarcino Hall, 901 Stange Road, Ames, IA 50011, United States. Email: tessneal@iastate.edu

Abstract

Objective: Across two experiments, we examined three cognitive biases (order effects, context effects, confirmatory bias) in licensed psychologists' diagnostic reasoning. ***Hypotheses:*** Our main prediction was that psychologist-participants would seek confirming versus disconfirming information after forming an initial diagnostic hypothesis, even given multiple opportunities to seek new information in the same case. We also expected that individual differences would affect diagnostic reasoning, such that psychologists with lower (vs. higher) cognitive reflection tendencies and larger (vs. smaller) bias blind spots would be more likely to demonstrate confirmatory bias. ***Method:*** In Study 1, we recruited 149 licensed psychologists (M = 18 years of experience; 44% women; 71% White) and exposed them to one of four randomly assigned vignettes that varied order effects (one set of symptoms in reversed orders) and context effects (court referral vs. employer referral). They rank ordered a list of four possible initial diagnostic hypotheses and received a piped follow-up choice of which of two pieces of information (confirmatory or disconfirmatory) they wanted to test their initial hypothesis. Study 2 (n = 131; M = 21 years of experience; 53% men; 68% White) replicated and extended Study 1, following the same procedure except offering three sequential choice opportunities. ***Results:*** Both studies found robust confirmatory information seeking: 92% sought confirmatory information in Study 1, and confirmation persisted across three opportunities in Study 2 (90%, 84%, 77%), although it lowered with each opportunity (generalized logistic mixed regression model), $F(2, 378) = 3.85$, $p = .02$, $\eta_p^2 = .02$. ***Conclusion:*** These findings expand a growing body of research on bias in expert judgment. Specifically, psychologists may engage in robust confirmation bias in the process of forming diagnoses. Although further research is needed on bias and its impact on

accuracy, psychologists may need to take steps to reduce confirmatory reasoning processes, such as documenting evidence for and against each decision element.

Keywords: order effects, context effects, cognitive bias, confirmation bias, diagnostic reasoning

Public Significance Statement

These studies revealed strong evidence that when psychologists form an initial diagnosis, they seek confirming information in an effort to determine whether their diagnosis is correct (rather than seeking information to determine whether their diagnosis is incorrect). Confirmation bias is a problem because overlooking disconfirming information can lead to misdiagnosis, failure to receive proper treatment, unjust sanctions, and more. Psychologists should take steps to reduce confirmatory bias in their diagnostic reasoning, such as carefully considering and documenting evidence both for and against each element of their decision process.

Confirmatory Information Seeking Is Robust in Psychologists' Diagnostic Reasoning

Human cognitive processing abilities are extraordinary, but we can make predictable and systematic errors because of the design of our cognitive machinery (e.g., Hoffrage & Gigerenzer, 2004; Kahneman & Klein, 2009). Decades of scholarship speak to the ways in which various psychological biases affect people's reasoning processes. For example, people tend to seek and rely on information that confirms their initial hypothesis rather than seeking potentially disconfirming information about that initial hunch. This behavior is called confirmatory bias and can lead to error (Nickerson, 1998). The modern scientific method evolved in part to combat the powerful confirmatory bias in hypothesis testing (Popper, 1959), yet evidence of confirmation bias persists in many contexts.

We sought to answer broad basic questions about experts' susceptibility to various biases in professional judgments and a narrower applied research question with direct implications for clinicians who work with people in the legal system. To do so, we conducted two experiments to investigate the effects of specific hypothesized biases (i.e., order effects, context effects, and confirmation bias) and individual differences (i.e., cognitive reflection abilities, size of bias blind spot) on psychologists' diagnostic reasoning processes.

Psychologists' Decision Task in Diagnosing Mental Disorders

One of the primary tasks of psychologists is to diagnose and treat mental disorders. Diagnosis involves a process of evaluation, including collecting information about symptoms, impairment of functioning, and level of distress, and then integrating that information into a diagnostic judgment that best captures the person's symptoms and prognosis (Frances & Widiger, 2012). The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., text rev.; *DSM-5-TR*; American Psychiatric Association, 2022) is the most recent version of a standardized

process and guide to aid clinical diagnostic judgment. It provides a common language and structure for clinicians and researchers and contains a coding system for insurance, administrative, and statistical purposes (Frances & Widiger, 2012). Despite the structure that the *DSM* helps to provide for diagnostic processes and communication, there remain significant challenges (for an overview, see Frances & Widiger, 2012). One of the major challenges is that it is difficult to identify ground truth for psychological diagnosis and thus to know whether and when a diagnosis is accurate.

Sources of information to inform a psychological evaluation can come from interviewing the evaluee; psychometric testing data; speaking with collateral sources such as family members, employers, medical providers, and teachers; and biological tests, among other sources (for the most common sources of information sought by psychological practitioners in the context of forensic evaluation, see Neal & Grisso, 2014a). There is little standardization between psychologists in their psychological assessment process as relevant to legal contexts (Neal, Martire, et al., 2022), with wide variation in the approaches that psychologists use and the data they seek to inform their expert judgments (e.g., Neal & Grisso, 2014a; Neal et al., 2019).

In expert judgment tasks characterized by greater validity, predictability, and objectivity, the scope for error and bias to emerge is limited (Cassidy & Buede, 2009; Kahneman & Klein, 2009; National Research Council, 2009; Shanteau, 1992). Conversely, in tasks marked by lower validity, predictability, and subjectivity, there is greater latitude for error and bias to develop and impact expert judgments. Several conditions under which clinicians make decisions increase the likelihood of bias, such as time pressure, integrating disparate information from various sources, little critical feedback, and framing and contextual effects (Acklin et al., 2015; Heilbrun & Brooks, 2010; Neal & Grisso, 2014b; Zapf & Dror, 2017).

One specific area in which diagnostic reasoning has evinced problems is in the differential diagnosis of antisocial personality disorder (APD). APD is one of 10 types of personality disorders specified in the *DSM-5-TR* (American Psychiatric Association, 2022). Personality disorders are long-standing patterns of behavior and subjective experiences of the world that differ from other people's behaviors and experiences of the world. APD in particular involves a long-standing pattern of disregard for and violation of the rights of other people (American Psychiatric Association, 2022). This diagnosis has become more controversial over the past few decades as information about challenges relating to validity and reliability of the diagnosis has emerged (for further discussion of diagnostic challenges, see Kotov et al., 2017).

Some of the problems associated with diagnosing APD include shifting diagnostic criteria (American Psychiatric Association, 1968, 1980, 1994, 2013), the number of symptom combinations that could result in an APD diagnosis (Rogers et al., 1994), and diagnostic overlap with substance use disorders (Cunningham & Reidy, 1998; Forbes et al., 2024). In fact, up to two thirds of people who qualify for a substance use disorder may also meet diagnostic criteria for APD (Brooner et al., 1992; Regier et al., 1990; Skodol et al., 1999). Given that the diagnostic criteria overlap, it is possible that clinicians are not differentiating well between the disorders (see Forbes et al., 2024). Resultant errors may have serious and long-lasting implications for people who are misdiagnosed. Accurate diagnoses help inform what to treat and how to treat; misdiagnosis can lead not only to unjust legal outcomes but also to iatrogenic treatment effects by providing unnecessary and perhaps even harmful treatment to people who do not need it.

Psychologists' Diagnostic Decision Task in Legal and Justice Settings

Given the reliance of judges on the opinion of forensic mental health evaluations and the potential impact of psycholegal opinions in various domains (Zapf et al., 2004), reaching an

accurate opinion is vital for justice. Many jurisdictions require the evaluator to identify a mental health diagnosis and subsequently associate the diagnosis to the elements of the psycholegal question (Melton et al., 2017), as a part of forensic mental health referral questions. For example, in assessing criminal responsibility, evaluators assess for a psychiatric diagnosis and whether it directly impacted the defendant's ability to understand the wrongfulness of the criminal behavior or to resist the behavior. Therefore, the diagnosis assigned to the examinee often underlies the psycholegal opinion. Errors in formulating a diagnostic picture of a defendant can substantially alter the psycholegal opinion, which could have significant legal consequences. For example, erroneously attributing a defendant's symptoms to a personality disorder or substance use (rather than a psychotic illness, for example) could result in an inaccurate opinion that a defendant is ineligible for an insanity defense in some jurisdictions and, thus, criminally responsible and deserving of punishment.

Regarding the specific challenge of the differential diagnosis of APD and substance use disorder, APD and substance use disorder can be comorbid and are not mutually exclusive. However, it is worth trying to distinguish between the two conditions, given the potentially different clinical and forensic implications of the two diagnoses. For example, if someone is showing traits consistent with APD only in the context of active substance use, that would suggest that the symptoms are not traits of a personality disorder but, rather, symptoms of or functional impairment resulting from a substance use disorder. This distinction is important for accurate treatment from a clinical perspective, but it is also critically important in the legal system to inform accurate risk assessment for bail determinations; accurate prognostic and risk assessment for probation, parole, and sentencing outcomes; accurate parental capacity and

functioning assessment for child custody, visitation rights, and child protection determinations; and many other issues.

Perhaps most troubling in light of the potential rate of error is the stigma associated with the APD diagnosis and its aggravating effect on sentencing considerations, including creating expectations that no rehabilitation is possible, that recidivism is inevitable, and that mitigating circumstances are irrelevant (Cunningham & Reidy, 1998). The negative impacts of this diagnostic label were demonstrated in two notorious U.S. Supreme Court cases that subsequently shaped legal policy and clinical practice regarding information about clinical assessments of violence potential: *Estelle v. Smith* (1981) and *Barefoot v. Estelle* (1983). Thus, the potential consequences of receiving the APD diagnosis are great and are unnecessarily borne by erroneously diagnosed people, especially as they intersect with the legal system.

Bias and How It Affects Human Judgment

A prevailing view among scholars is that bias is an inadvertent outcome arising from typical human cognitive processes. A widely accepted definition of bias is that it is any systematic factor influencing judgment aside from the truth (West & Kenny, 2011). Although there is general scholarly agreement about this concept of bias and how it operates, there are different perspectives about the functional value of bias in judgment (see Kahneman & Klein, 2009).

Bias can arise from at least 200 different cognitive sources (Neal, Lienert, et al., 2022). We might hope and expect experts to be protected against predictable biases, but the scientific evidence suggests that experts are human and susceptible to bias just like the rest of us (Murrie et al., 2013; Neal & Pronin, 2023). Expertise is real and improves many skills related to discrete domains (Cassidy & Buede, 2009; Kahneman & Klein, 2009; Shanteau, 1992), but experts are

subject to the bounded rationality inherent in the limitations of the human brain's design (Simon, 1956). Within the confines of our cognitive capabilities, people—including experts—may resort to cognitive shortcuts or simplifying strategies to manage cognitive load (Kahneman, 2011). Thus, expertise may not necessarily protect against the kinds of systematic biases that affect human judgment (Growns & Neal, 2024).

Neal, Lienert, and colleagues (2022) synthesized a large body of evidence from diverse fields and theories about the circumstances under which people (including experts) are protected against bias and when they are particularly vulnerable to it. They introduced, on the basis of that synthesis, a comprehensive framework for organized, testable predictions about specific manifestations of bias that are likely to emerge as well as strategies to mitigate these biases. This descriptive model outlines how cognitive biases impact human judgment across two interconnected continuous dimensions: depth of cognitive processing (ranging from low to high) and susceptibility to bias (ranging from low to high; see Figure 1).

[Insert Figure 1 about here]

In specific scenarios, relying on low-effort intuitive cognitive processes may lead to bias (lower left side of Figure 1) but offer protection against bias in other foreseeable situations (upper left side of Figure 1). Similarly, in predictable conditions, engaging in high-effort deliberative cognitive processes may heighten the risk of bias (lower right side of Figure 1) but could mitigate it in alternative situations (upper right side of Figure 1). This bias model aligns with contemporary cognitive frameworks that depict human decision making as existing along a continuum (Hammond, 1996; Kruglanski, 2013; Kruglanski & Gigerenzer, 2011) or even across multiple continua (Varga & Hamburger, 2014).

Some Specific Types of Bias Relevant to This Project: Order Effects, Contextual Effects, and Confirmation Bias

Order effects refer to the impact that the sequence of information has on people's perception and judgments (Asch, 1946; Hogarth & Einhorn, 1992). That is, the same information can significantly influence judgments and decisions differently based on the order in which those pieces of information are encountered. One robust order effect is primacy, where information encountered earlier tends to have more enduring impact on judgment than information encountered later (this is especially true for judgments that are formed only after all information is obtained; Hogarth & Einhorn, 1992). This cognitive bias suggests that people give disproportionate weight to the initial information they receive, shaping their first impressions and ultimately overall decisions. Conversely, the recency effect posits that information encountered most recently may exert a stronger influence on judgments because of salience (this effect tends to emerge when people form their judgments in a step-by-step fashion explicitly after each item of information is obtained; Hogarth & Einhorn, 1992). Order effects such as these showcase the malleability of human judgment, illuminating how cognitive processes are affected by the sequence in which information is presented.

Beyond order effects, context effects have been extensively documented in human judgment. People can arrive at different conclusions using the same information, depending on how it is presented or the context in which it is presented (Tversky & Kahneman, 1981). Sometimes, contextual information can be relevant to a given decision task, but other times, contextual information carries biasing power without much relevance. When contextual information is biasing yet available in a decision environment, it increases the risk of contextual bias—a psychological process whereby people draw on task-irrelevant contextual information

from their environment when making decisions. Contextual effects have been observed across various forensic disciplines, including fingerprint examination (Fraser-Mackenzie et al., 2013; Langenburg et al., 2009; Stevenage & Bennett, 2017), firearms examination (Mattijssen et al., 2020), forensic document examination (Miller, 1984), footwear examination (Sneyd et al., 2020), bitemark examination (Osborne et al., 2014), blood spatter analysis (Osborne et al., 2014, 2016), and even in what is regarded as the "gold standard" in forensic science—DNA analysis (Dror & Hampikian, 2011).

A third specific type of bias that is well documented in human judgment is confirmation bias, which represents a collection of inclinations to search for or interpret evidence in ways that align with preexisting beliefs, expectations, or hypotheses (Klayman & Ha, 1987; Nickerson, 1998). Confirmation bias is ubiquitous in human judgment (see Elstein et al., 1978; LeBlanc et al., 2002; Schulz-Hardt et al., 2000), even in the absence of specific motivational factors (Fischhoff & Beyth-Marom, 1983; MacCoun, 1998). Even expert judgments can be vulnerable, such as in intelligence analysis (Cook & Smallman, 2008), criminal investigations (Ask & Granhag, 2005), medical reasoning (e.g., Drew et al., 2013; Elstein et al., 1978), and scientific reasoning (MacCoun, 1998).

Confirmation bias is important to study and ultimately to mitigate in clinical contexts because it can lead to error, misdiagnosis, failure to receive proper treatment, unjust sanctions, and more. Nickerson (1998) wrote, "The guidance that a hypothesis in hand represents for further information gathering can function as a constraint, decreasing the likelihood that one will consider an alternative hypothesis if the one in hand is not correct" (p. 193). Previous research has shown that a faulty medical diagnosis can result from incorrect initial hypotheses (e.g., Barrows et al., 1978; Ludolph & Schulz, 2018), and an initial hypothesis can bias interpretation

of subsequent data, including in studies with diagnostic relevance (e.g., Elstein et al., 1978). Importantly, clinicians' diagnostic reasoning can be studied as a domain-specific instantiation of more general cognitive processes, and new understandings generated from these kinds of studies can add to basic cognitive science in judgment and decision making (Marsh et al., 2018).

The Current Project: Investigating Psychologists' Diagnostic Reasoning

Despite increasing attention to bias and error in expert human judgment in legal contexts (e.g., National Research Council, 2009; President's Council of Advisors on Science and Technology, 2016), few experimental studies have measured bias in psychologists' judgments (Neal, Lienert, et al., 2022). Better understanding the extent to which, and circumstances under which, experts are susceptible to versus protected from bias is useful for informing debiasing interventions. Likewise, better understanding the factors that differ between psychologists that could make them more versus less susceptible to bias also has potential to inform debiasing interventions. Ultimately, fair, just, accurate, and unbiased professional judgments are the goal (American Psychological Association, 2017). Across two experimental vignette studies, we examined the extent to which three different cognitive biases affect psychologists' diagnostic reasoning as well as how individual differences in cognitive reflection and the bias blind spot affect diagnostic reasoning. Study 2 replicated and extended Study 1. Materials, data, and code are available on the Open Science Framework (https://osf.io/678cs/).

Hypothesis 1: Order Effects

We expected diagnostic judgments to be influenced by the order of information encountered, such that psychologists randomly assigned to Symptom Order 1 would be more likely to reach an initial diagnosis of APD than substance use disorder, whereas the symptoms in reverse order would lead to more substance use disorder diagnoses than APD diagnoses. The

symptoms described in the scenarios could reasonably fit the diagnostic criteria for either disorder. Both orders listed the same eight symptoms, but Symptom Order 1 listed the strongest APD-consistent symptom first and vice versa.

Hypothesis 2: Context Effects

We expected diagnostic judgments to be influenced by context effects, such that psychologists randomly assigned to Vignette 1 (involving a court referral for an assessment of a defendant's behavior prior to sentencing) would be more likely to reach an initial diagnosis of APD than substance use disorder, whereas psychologists randomly assigned to Vignette 2 (involving a workplace referral to assess an employee with the same behaviors on the basis of concerns that the behavior would impact performance at work) would be more likely to diagnose substance use disorder than APD.

Hypothesis 3: Confirmation Bias

After reaching an initial diagnostic hypothesis, participants were asked to select which of two pieces of information they would like to test their initial hypothesis. We expected that clinicians would be more likely to choose the confirmatory than disconfirmatory information (i.e., they would seek information to try to prove rather than disprove their hypothesis), even given multiple opportunities to seek disconfirming information.

Hypothesis 4a: Individual Differences

Cognitive reflection is the ability to reflect on a question and resist the first "heuristic" response that comes to mind, instead engaging in deliberative thought to reach an answer (Frederick, 2005). We predicted that clinicians with higher cognitive reflection tendencies would be less likely to engage in confirmatory bias. We also predicted that cognitive reflection would be inversely related to the size of the bias blind spot.

Hypothesis 4b: Individual Differences

The bias blind spot, a metacognitive bias, refers to the seemingly universal human tendency to recognize the impact of bias on other individuals more readily than on ourselves (i.e., to be blind to our own biases but not to those of other people; Pronin et al., 2002; Pronin & Hazel, 2023). We hypothesized that the size of participants' bias blind spot (discrepancy between self- and other-rating on susceptibility to bias) would be positively related to confirmatory bias and negatively related to cognitive reflection.

Study 1

Method

We asked a national sample of licensed psychologists to consider and provide diagnostic hypotheses for and answer questions about vignettes of people presenting with a set of psychological symptoms. We designed the vignettes to investigate particular cognitive biases, including (a) *order effects* based on a list of symptoms, (b) *context effects* based on referral source, and (c) *confirmation bias*. We randomly assigned each clinician to one of four vignettes, in which symptom order (1 vs. 2) and referral context (court vs. employer) were experimentally manipulated and a question followed that focused on confirmation bias.

Participants

An a priori power analysis with an anticipated medium effect size and with a power level of 0.8 and $p < .05$ indicated that we needed a minimum of 91 participants (Cohen, 1992). To estimate our anticipated effect size, we relied on previous studies documenting the medium-to-large effects of people's susceptibility to cognitive biases such as the anchoring heuristic (West et al., 2012) and context effects (Murrie et al., 2013).

To sample our population of interest, we developed a large, national database of licensed psychologists with forensic interests (this two-paragraph description of the development of our database first appeared in the work by MacLean et al., 2019). Specifically, we used licensing databases maintained by each of the 50 states within the United States. We reviewed the available specialty search options for each state separately. The search options on each state licensing database differed, and we maintained a document detailing the search criteria used for each state. We selected the search options that were most relevant to forensic evaluators and compiled a national list (e.g., court-ordered evaluations, forensic evaluations, custody evaluations, disability determination, risk assessment, personal injury, criminal responsibility, juvenile delinquency, fitness for duty, capacity evaluation, court testimony). Our database included 2,221 licensed psychologists across the nation with forensic interests. After we had compiled the list of names, a team of research assistants found contact information for most of these psychologists, including professional mailing addresses, professional email addresses, and professional websites.

Our goal in adopting this sampling procedure—rather than more common convenience strategies of posting study invitations to professional electronic mailing lists or sending invitations to members of professional groups (e.g., American Psychology-Law Society)—was to better sample the population of practicing psychologists with forensic interests. Our effortful sampling method affords two significant improvements over more common and convenient methods of sampling. First, the new database that we created represents the population of licensed psychologists with forensic interests across the entire nation, rather than only those who sought membership in a professional society. Second, the database was sufficiently large that we could then randomly select participants from the population.

We used a random number generator to randomly select participants from the overall list and then contacted them through email with an invitation to participate in this study. If we did not have an email address for the randomly selected participant, we replaced the random selection with a new randomly generated selection. We sent 727 email invitations but discarded 35 because of an incorrect email address or because the recipient replied that they lacked forensic experience. We collected data from 149 licensed psychologists across the United States (22% response rate based on the 692 presumably valid email invitations; 117 completed all questions, but we report all the data available as applicable to each analysis).

Both men (n = 50, 33.6% of the sample) and women (n = 65, 43.6% of the sample) participated (22.8% did not indicate their gender), with a mean age of 53.30 years (SD = 13.77). Participants were predominantly White (71.1% of the sample; 2.0% Latinx, 0.7% American Indian/Native American, 0.7% Asian/Asian American, 0.7% African American/Black, 2.0% other; 22.8% of the sample did not indicate their race). Participants reported a mean of 18.44 years of experience (SD = 11.45). Participants worked in private practice (n = 86, 57.7% of the sample), institutions or agencies (8.1%; e.g., hospital, department of health, Veterans Affairs hospital, court clinic, prison), university settings (6.7%), and other settings or more than one setting (4.0%; 23.5% of the sample did not specify employment). Regarding how they spend their work time, most of their monthly time was in clinical (nonforensic) practice (M = 49.59%, SD = 32.88%), followed by forensic practice (M = 30.32%, SD = 31.68%), administrative duties (M = 7.68%, SD = 10.25%), consultation (M = 4.20%, SD = 11.05%), teaching (M = 4.90%, SD = 10.95%), and research (M = 3.31%, SD = 7.99%).

Procedure

We invited participants by email, with a link to the 20-min survey in Qualtrics. After completion, we compensated them with a $20 Visa gift card. After the informed consent process, participants provided diagnostic hypotheses for, and answered questions about, one of four randomly assigned vignettes of people presenting with a set of symptoms. The vignettes were about 300 words, designed so the symptoms described could reasonably fit the diagnostic criteria for either APD or alcohol use disorder. The initial diagnostic question asked participants to rank order a list of four possible diagnostic hypotheses "in order of likelihood that this person may meet *DSM-5* diagnostic criteria for each" (the initial options were the same across vignettes and presented in randomized order). From there, they received a piped follow-up question linked to the diagnostic hypothesis that they rank ordered first.

The follow-up question asked, "Now, based on your primary diagnostic hypothesis that Mr. G meets criteria for x, what piece of information would you want first in order to effectively test your primary diagnostic hypothesis?" with a choice between two types of information: one that might confirm their initial hypothesis, and one that might disconfirm it (presented in randomized order; see Figure 2). For example, participants who rank ordered APD as their first diagnostic hypothesis had the choice between "Has Mr. G shown a pervasive pattern of disregard for and violation of the rights of others since at least 15 years of age?" (*confirmatory*, as it is a criterion from the *DSM-5-TR* for diagnosing APD; American Psychiatric Association, 2022) and "Does Mr. G have a substance use disorder that could explain his symptoms?" (*disconfirmatory*). Participants who rank ordered alcohol use disorder first had the choice between "Does Mr. G show evidence of alcohol tolerance and withdrawal?" (*confirmatory*, from the *DSM-5-TR*; American Psychiatric Association, 2022) and "Does Mr. G have a personality disorder that could explain his symptoms?" (*disconfirmatory*). The options for intellectual disability and dissociative

identity disorder followed the same logic but were much less relevant to the content of the vignette, and few psychologists selected either of these latter diagnoses as their initial hypotheses. These options functioned as manipulation checks.

[Insert Figure 2 about here]

We set up the language of the choices such that the disconfirmatory option would have the highest probability of yielding the most diagnostic value, as it would be most likely to critically falsify the hypothesis (Klayman & Ha, 1987). In other words, it could reveal to the psychologist that they were on the wrong diagnostic track with just a single test. Selecting the confirmatory option would yield less diagnostic value because regardless of the answer, further information would be required to either rule out or rule in the initial hypothesis (see Figure 2; see materials on the Open Science Framework).

Measures

After the diagnostic question and the piped follow-up, participants responded to the Cultural Cognition Scale, Cognitive Reflection Task (CRT; Frederick, 2005), and bias blind spot questions. They then responded to a series of questions about awareness of various biases and strategies to reduce bias (results from these survey questions were reported by MacLean et al., 2019, and are not discussed here). Demographics followed. Note that the Cultural Cognition Scale (Kahan et al., 2011) did not perform well and thus is not reported on further in this study (although the analytic code available on the Open Science Framework includes further information about cultural cognition theory, the scale itself, and how it performed in our analyses in this study for interested readers).

Cognitive Reflection Task. The CRT is a brief questionnaire that measures abilities to inhibit initial instinctive but incorrect responses in favor of deliberate and reflective correct

responses. The measure includes three right/wrong items, with a maximum score of 3 for reflective and correct responses. For example, "A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents." Given the items' design to solicit an instinctive but incorrect response, higher scores are characteristic of individuals with more advanced cognitive reflection abilities. The average number of items correct on the CRT in this sample was 1.10 (SD = 1.18, α = .79).

Bias Blind Spot. We asked participants to rate, on a scale ranging from 1 (*never*) to 9 (*always*), the extent to which their own forensic work is influenced by bias (M = 4.80, SD = 2.06) as well as the extent to which work by other forensic psychologists is influenced by bias (M = 4.73, SD = 2.11). These questions measured participants' bias blind spot, the tendency to deny personal bias even while recognizing it in other people (Pronin et al., 2002). We calculated bias blind spot by subtracting self-rating from other-rating, paired-samples $t(107) = -0.32$, p = .75, Cohen's d = –0.03, 95% confidence interval (CI) = [–0.22, 0.16].

Results

Contrary to Hypothesis 1, the order of symptoms encountered in the vignette did not systematically affect psychologists' initial diagnostic judgments, $\chi^2(1, 142) = 1.57$, p = .21, OR = 1.07, 95% CI [0.53, 2.14]. Nor did the contextual information about referral source (court vs. work supervisor) systematically affect psychologists' initial diagnostic judgments (Hypothesis 2), $\chi^2(1, 142) = 0.03$, p = .86, OR = 1.57, 95% CI [0.77, 3.19] (see Table 1). However, consistent with Hypothesis 3, after they made their initial hypothesis, psychologists overwhelmingly sought confirmatory information (92%; n = 130) rather than disconfirmatory information (8%; n = 11) about that initial hypothesis.

[Insert Table 1 about here]

Hypothesis 4 involved the degree to which individual differences between psychologists would relate to their likelihood of engaging in confirmatory reasoning. First, to test the prediction that clinicians with higher cognitive reflection tendencies would be less likely to engage in confirmatory bias (Hypothesis 4a), we conducted a logistic regression with information choice as the dependent variable (confirming vs. disconfirming) and the three-item CRT total as the predictor. Cognitive reflection had a statistically and theoretically significant association with confirmation bias in the predicted direction. Each unit increase on the three-item CRT (representing higher cognitive reflection tendencies) halved the odds of seeking confirmatory information, $B = –0.59$, $Wald(1) = 4.65$, $p = .03$, $OR = 0.55$, 95% CI [0.32, 0.95].

Second, to test the prediction that clinicians with larger bias blind spots would be more likely to engage in confirmatory bias (Hypothesis 4b), we conducted a logistic regression with information choice as the dependent variable (confirming vs. disconfirming) and the size of a psychologist's bias blind spot (operationalized as the discrepancy between self- and other-ratings of bias) as the predictor. Contrary to expectations, size of bias blind spot did not systematically relate to the behavior of seeking confirmatory information, $B = –0.04$, $Wald(1) = 0.06$, $p = .80$, $OR = 0.96$, 95% CI [0.70, 1.31].

Consistent with an expected negative correlation between the two individual difference measures (Hypotheses 4a and 4b), cognitive reflection tendencies inversely related to the size of the bias blind spot ($r = –.27$, $p = .007$). This means that the lower psychologists scored on the cognitive reflection items, the larger the difference between how biased they believed they were compared with their peers (or, stated another way, the higher psychologists scored on the cognitive reflection items, the closer their estimates were between how susceptible to bias they believed they were compared with how susceptible to bias they believed their peers were).

Discussion

Although order and contextual effects did not emerge in this simple vignette study of psychologists' diagnostic judgments, we did observe powerful confirmatory information-seeking behavior. It may be that biases from order and contextual effects would emerge in more ecologically valid, complicated tasks, whereas this one was straightforward and brief. We were left with an open question about whether psychologists might start with seeking confirmatory information but then switch to include disconfirmation in their reasoning the further they went along in their judgment task. To understand how far confirmation might persist beyond the first piece of information, we replicated and extended this study, giving psychologists multiple information-seeking opportunities. People may be more confirmatory seeking when information comes in sequentially rather than simultaneously (Jonas et al., 2001). But, in a sequential diagnostic decision environment (which has some real-world applicability), we wondered how far they would stay in confirmatory mode (how many pieces of information they would seek in a confirmatory manner).

Study 2

Method

Following the same general procedure as Study 1, we asked psychologist-participants to rank order a list of four possible initial diagnostic hypotheses "in order of likelihood that this person may meet *DSM-5* diagnostic criteria for each." They received, on the basis of their response, three follow-up questions linked to the diagnostic hypothesis that they rank ordered first. The follow-up questions presented a series of choices about further information they would want to know to effectively test their initial diagnostic hypothesis. Each of the three questions provided them with a choice between two types of information: one that might confirm their

initial hypothesis, and one that might disconfirm it (see materials on the Open Science Framework).

Participants

To complement the email recruitment strategy for Study 1 and invite new participants who were never contacted for Study 1, we sent invitations to participate in Study 2 by postal mail. Following the recruitment method of Neal and Brodsky (2016), we relied on several strategies to increase participation likelihood (i.e., postal survey, university sponsorship, green paper, first-class postage, one-dollar bill, follow-up postcard; Fox et al., 1988; King & Vaughan, 2004). Inside the mailed envelope, we included a recruitment letter describing the study and the web address where participants could go to engage in it, as we hosted data collection on Qualtrics. We randomly selected people to invite from the database described in Study 1. If we did not have a postal address for the randomly selected participant, we replaced the random selection with a new randomly generated selection. We mailed 885 individual invitations to participate; 127 were returned as undeliverable.

We collected data from 131 licensed psychologists across the United States (17% response rate based on the 758 presumably valid addresses; 100 completed all questions, but we report all the data available as applicable to each analysis). Both men (n = 69, 52.6% of the sample) and women (n = 31, 23.7% of the sample) participated (23.7% did not indicate their gender), with a mean age of 59.43 years (SD = 10.97). Participants were predominantly White (67.9% of the sample; 3.1% Latinx, 0.8% African American/Black, 0.8% Asian/Asian American, 3.8% other; 23.6% of the sample did not indicate their race). The participants reported a mean of 21.06 years of experience (SD = 11.58). Participants worked in private practice (52.6% of the sample), institutions or agencies (13.0%; e.g., hospital, department of health, Veterans Affairs

hospital, court clinic, prison), university settings (6.1%), and other settings or more than one setting (3.1%; 25.2% of the sample did not specify employment). Regarding how they spend their work time, most of their monthly time was in clinical (nonforensic) practice (M = 48.50%, SD = 30.87%), followed by forensic practice (M = 29.47%, SD = 29.86%), administrative duties (M = 8.86%, SD = 11.20%), consultation (M = 6.37%, SD = 13.33%), teaching (M = 5.34%, SD = 13.18%), and research (M = 1.46%, SD = 5.30%).

Measures

After answering the diagnostic question and the three piped follow-up questions, participants responded to the Cultural Cognition Scale, CRT, and the bias blind spot questions. Demographics followed. As in Study 1, we do not report further on the Cultural Cognition Scale (Kahan et al., 2011), although the analytic code available on the Open Science Framework includes further information about how it performed in this study.

Cognitive Reflection Task. The same three-item CRT as in Study 1 was used in this study (Frederick, 2005). In addition, we used four more CRT items developed by Toplak and colleagues (2014) as well as four more CRT items developed by Thomson and Oppenheimer (2016). The average number of items correct on the three-item CRT in this sample was 1.35 (SD = 1.19, α = .76). All 11 CRT items together had an average number correct of 6.69 (SD = 2.42, α = .67).

Bias Blind Spot. Participants rated, on scales ranging from 0% to 100%, the extent to which their own forensic work is influenced by bias (M = 28.44%, SD = 19.01%) as well as the extent to which work by other forensic psychologists is influenced by bias (M = 33.61%, SD = 20.04%). As in Study 1, we calculated the bias blind spot by subtracting self-rating from other-rating (M = 5.17%, SD = 9.49%), paired-samples $t(113) = 5.81$, $p < .001$, Cohen's d = 0.54, 95%

CI [0.35, 0.74]. Beyond the questions about abstract bias, we also asked three questions about one's own and other individuals' susceptibility to three specific biases in this particular study—that is, contextual effects (M_{self} = 34.10, SD_{self} = 22.82; M_{other} = 34.12, SD_{other} = 21.09), confirmation (M_{self} = 31.12, SD_{self} = 21.54; M_{other} = 34.51, SD_{other} = 21.19), and drawing different conclusions from the same information depending on how the information is presented (M_{self} = 32.63, SD_{self} = 21.71; M_{other} = 34.90, SD_{other} = 20.66). Paired-samples *t* tests revealed significant differences between self-other ratings for confirmation bias, $t(113) = 2.83$, $p = .005$, Cohen's d = 0.26, 95% CI [0.08, 0.45], but not for how information is presented, $t(113) = 1.71$, $p = .09$, Cohen's $d = 0.16$, 95% CI [–0.03, 0.35], or contextual effects, $t(113) = 0.02$, $p = .98$, Cohen's d = 0.002, 95% CI [–0.18, 0.19].

Results

Diverging from Study 1 but consistent with Hypothesis 1, order of symptoms encountered in the vignette systematically affected psychologists' initial diagnostic judgments, $\chi^2(1, 124) = 5.20$, $p = .02$, $OR = 2.38$, 95% CI [1.12, 5.06] (see Table 1). These findings mean that psychologists exposed to the evaluee "having a short fuse" as the first of the eight symptoms described were 2.38 times more likely to reach an initial diagnostic judgment of APD rather than alcohol use disorder compared with psychologists exposed to alcohol use as the first of the same eight symptoms. Replicating Study 1 but contrary to Hypothesis 2, the contextual information about referral source (court vs. work supervisor) did not systematically affect psychologists' initial diagnostic judgments, $\chi^2(1, 124) = 0.99$, $p = .32$, $OR = 0.69$, 95% CI [0.33, 1.44] (see Table 1).

Replicating Study 1 and confirming Hypothesis 3, at the first information opportunity, psychologists overwhelmingly sought confirmatory information (90%; n = 118) rather than

disconfirmatory information (10%; $n = 13$) about their initial hypothesis. Extending Study 1 and confirming Hypothesis 3, confirmatory information seeking persists: At the second choice, 84% chose confirmatory, and at the third, 77% chose confirmatory (see Figure 3). Overall, 67% chose confirmatory information at all three choice points.

[Insert Figure 3 about here]

Although confirmatory information-seeking behavior remained high across the three decision opportunities, it lowered with additional opportunities—generalized logistic mixed regression model, $F(2, 378) = 3.85$, $p = .02$ (see Table 2 and Figure 3). Specifically, although the generalized logistic mixed regression slope for Trials 1 and 2 did not differ, the slope for Trial 3 was negative and significant compared with Trial 1, indicating that by the third choice opportunity, participants became less likely to choose confirming information (see Table 2). Of note, there was no significant random effect of individual participants in the model, suggesting that the pattern of confirmatory information seeking applies similarly across the experts (see Table 2).

[Insert Table 2 about here]

Contrary to Hypothesis 4a, cognitive reflection did not relate to number of confirmatory choices (three-item CRT: $r = .08$, $p = .36$; 11-item CRT: $r = .04$, $p = .73$). Neither did cognitive reflection relate to any of the three confirmatory choice opportunities individually (contrary to Study 1 and Hypothesis 4a). Contrary to the findings of Study 1, cognitive reflection tendencies did not relate to the bias blind spot in general; however, in partial support of Hypotheses 4a and 4b, some of the specific bias blind spots inversely related to cognitive reflection (see Table 3).

[Insert Table 3 about here]

The size of a psychologist's bias blind spot did not systematically relate to number of confirmatory choices (r = .10, p = .29). Contrary to Hypothesis 4b, a series of logistic regressions did not uncover any relationship in this study between the general bias blind spot or any of the specific bias blind spots and the behavior of seeking confirming information.

Discussion

These findings further demonstrate robust evidence of confirmatory information seeking in diagnostic judgments in a national sample of licensed psychologists in the United States. Confirmation persists beyond just the initial hypothesis testing opportunity: This study shows that confirmatory behavior persists for at least three sequential testing opportunities (although it does decrease with additional opportunities). We need further research to understand how far confirmation bias persists (beyond three initial tests) and how this pattern may differ when information becomes available to clinicians before they have to make additional choices for what information to seek. If confirmatory information seeking leads to biased judgments, we need further research on how to reduce confirmation bias in expert judgment.

We again did not observe context effects, although we did observe an effect of symptom order on diagnostic judgments, such that the first symptoms encountered affected initial hypotheses more strongly than symptoms encountered later. Although conceptually, the vignettes and experimental design were replicated across Studies 1 and 2, the language of the vignettes was slightly different in Study 1 compared with Study 2. This means that Study 2 was a conceptual replication and extension of Study 1 but was not an exact replication.

In addition, in Study 2, the language of the symptoms was not exactly the same in the reversed orders: The words "heavy daily use" of alcohol and marijuana were included in Symptom Order 1 where they appeared first in the list of symptoms (which was the language

used in both Symptom Orders 1 and 2 in Study 1), but for Symptom Order 2 in Study 2, the phrase "a pattern of using" alcohol and marijuana was used. This difference in language is substantive, as heavy daily use is more consistent with a substance use disorder than a pattern might be, which should impact diagnostic judgments. Given that the phrasing changed in addition to the order, it is not clear that the order effect is entirely what drove the significant order finding in Study 2.

General Discussion

The current studies suggest that experts do not seem immune to confirmatory bias, at least in the psychological diagnostic judgment context of these particular studies. Importantly, the language of the choices available to psychologists in the critical tests of their reasoning processes reveals clear evidence of a positive test strategy that does indeed appear to reflect robust confirmation bias (see Klayman & Ha, 1987). Specifically, the language of the disconfirmatory options made clear that they carried the highest probability of diagnostically informative content. We set up the disconfirmatory options such that a single answer to those tests of the diagnostic hypothesis could critically falsify it (i.e., if the answer was "no," psychologists could know with one step that they were on the wrong diagnostic track). Selecting the confirmatory option would yield less diagnostic value because regardless of the answer, further information (i.e., more hypothesis-testing steps) would be required to either rule out or rule in the initial hypothesis. Testing options most likely to falsify a hypothesis are generally optimal (Klayman & Ha, 1987). Selecting the confirmatory choice was rational as it was related to the diagnostic task and would carry relevant information (and thus was not "pseudodiagnostic"; Crupi et al., 2009); however, it was less optimal, probabilistically, than the disconfirmatory choice.

The finding that psychologists first seek confirmatory evidence in support of their initial clinical hypothesis may be a concern but must be considered in the context of this study. In both Studies 1 and 2, participants provided an opportunity to seek additional information overwhelmingly requested information that would confirm their diagnosis. Study 2 demonstrated that this preference persisted across choice options such that by the third option to seek additional information, participants were quite likely to still be seeking confirmatory information. Although this finding does not preclude the likelihood that clinicians would seek disconfirming information through the course of their real-world clinical decision making, it does demonstrate a bias for preferring information that confirms their initial clinical hypotheses (and increases the risk of further premature commitment to an erroneous decision).

These studies did not provide the answers psychologists sought before they had to make additional choices for more information. Thus, they did not have a chance to update their priors before making decisions for subsequent information (see Lejarraga & Hertwig, 2021). It is possible that psychologists would adjust course and not continue seeking confirmatory information if they received information either consistent or inconsistent with that initial hypothesis, but these data do not yield insights to that possibility.

In addition, we examined a series of judgments that psychologists made sequentially, but psychologists may be less likely to engage in confirmation seeking when information is available simultaneously (see Jonas et al., 2001). Thus, the sequential mode that we examined here could be compared with a more simultaneous setup in future research. Furthermore, we need further research to understand what policies and procedures can mitigate these well-understood biases and what practical steps practitioners can take to reduce cognitive bias. Regardless, as diagnostic systems improve in psychology (e.g., potentially moving away from discrete categorical

diagnoses and instead toward more dimensional approaches; Conway et al., 2021; Lilienfeld & Treadway, 2016), the potential for bias in diagnostic reasoning may be reduced.

Reducing Bias in (Expert) Human Judgment

Neal, Lienert, and colleagues' (2022) descriptive model of bias in judgment (see Figure 1) offers theoretical insights into how various debiasing strategies might contribute to reducing bias in judgment. For instance, deliberative processes can be engaged to reduce bias as highlighted in the upper right side of the figure. Motivational strategies aim to engage deliberative processes and reduce errors in decision making by elevating the stakes, such as holding individuals accountable for their decisions. Accountability, involving the expectation of justifying a decision later on, could prompt experts to recognize flaws in their reasoning and thereby diminish the impact of various biases (Lerner & Tetlock, 1999). Or other deliberative cognitive strategies, such as the "consider-the-opposite" approach, could be employed to try to counteract biases. Asking oneself why an initial judgment might be incorrect and exploring alternative possibilities is a promising method for minimizing the effects of cognitive biases, including the hindsight bias (Arkes et al., 1988) and anchoring bias (Mussweiler et al., 2000). However, introspection is likely an ineffective debiasing method because the low-effort, automatic processes underlying many cognitive biases may not be accessible through introspection (Pronin & Kugler, 2007).

With regard to reducing bias through deliberative effort as reflected in the upper right area of Figure 1, we offer a specific suggestion for potentially reducing confirmation bias in expert judgment processes. Building on the "consider-the-opposite" approach described above but adapting it from a global practice that one might engage at the end of a diagnostic process to one that is engaged sequentially at each step of the process, we advise experts to carefully

consider and document evidence both for and against all of the elements of one's judgment process. This suggestion could theoretically reduce confirmation bias by prompting experts to constantly consider and pay attention to both evidence for and against their hypotheses as they emerge in the case; the early and consistent practice of seeking and considering both confirmatory and disconfirmatory information for each step of a judgment process may be key to reducing the effects of confirmatory reasoning. Future research may shed light on this proposition.

With sufficient repetition and motivation to automatize certain aspects, one can leverage expertise-enabled low-effort cognitive processing to diminish the effects of bias by overlearning bias correction, as illustrated in the upper left part of Figure 1. Alternatively, rather than reliance solely on individuals to consciously correct biases or invest significant effort to train themselves to do so automatically, the decision environment itself can be modified to mitigate the effects of bias (also reflected in the upper left part of Figure 1). In other words, bias can be managed by implementing bias-mitigating policies and procedures. An effective approach involves minimizing bias by ensuring that examiners are blinded to potentially biasing and task-irrelevant information that is not essential to their tasks. A related strategy is to regulate when and how examiners encounter such information, employing graded blinding procedures through information-management protocols (Krane et al., 2008; Quigley-McBride et al., 2022). With these methods, experts have restricted exposure to information, eliminating subjective freedom in determining the order in which they consider different pieces of information. Of course, these methods—recommended as reforms for forensic science procedures such as fingerprint analyses—may be more difficult to apply in the practice of clinical and forensic psychology, in which task characteristics vary substantially between different types of referral questions.

With regard to reducing bias without having to rely on clinicians' deliberative cognitive processes, we offer a suggestion for potentially reducing confirmation bias in expert judgment processes as reflected in the upper left area of Figure 1. Specifically, we suggest trying to limit one's exposure to only relevant and nonbiasing information during the evaluation process (e.g., for forensic psychologists, limiting interactions—especially early ones—with the referring party to scoping the referral question without information about the referring party's theory of the case or the arguments they would like to feature at trial).

Understanding how bias influences decision making empowers organizations and professions to proactively address workflow and organizational factors that may contribute to bias (Saposnik et al., 2016). Better understanding can inform strategies for mitigating bias toward enhancing justice and fairness in society at large (e.g., Forscher et al., 2019; Ludolph & Schulz, 2018; Sellier et al., 2019), with particular relevance to legal and administrative systems (e.g., Cooper & Meterko, 2019; Morewedge et al., 2015).

Implications for Forensic Practice

The results of these studies have important implications for educators training students in assessment as well as practicing mental health professionals. Regarding educators, these findings highlight the need for professional organizations and educators to enhance trainings (e.g., continuing education workshops, assessment courses) on evaluator bias. Educators training mental health professionals should incorporate strategies for reducing bias in clinical decision making throughout graduate student training and in continuing education offerings. Regarding practicing psychologists and other mental health professionals, these results highlight the need for practitioners to consider how they seek additional information.

Practically and specifically, how might these findings influence the practice of forensic mental health evaluation? After all, many requests for forensic evaluations are framed as hypotheses: Does this criminal defendant meet the criteria for legal insanity? Did this civil litigant suffer emotional injury from a defendant's behavior? Particularly when requested by one adversarial party, these requests are often accompanied by information that supports the hypothesis (e.g., accounts of psychiatric symptoms, complaints of harms suffered), a primacy effect that may strengthen an initial hypothesis. But certainly, evaluators must seek information in a sequential manner to explore many hypotheses. For example, if a defendant manifested serious psychiatric symptoms at the time of the offense, the evaluator must further explore whether those meet the additional narrow criteria for legal insanity. If a civil litigant manifests symptoms of emotional injury, the evaluator must further explore whether those are attributable to the alleged tort (vs. another cause). Evaluators considering a diagnosis must consider every relevant criterion (as in this study) before assigning the diagnosis.

The challenge for evaluators, of course, is to pursue relevant disconfirming information as vigorously as—if not more vigorously than—they pursue relevant confirming information. We suggest, as a starting point, that training programs and evaluation guidelines emphasize the "ubiquitous phenomena" and "many guises" of confirmation bias (Nickerson, 1998), the ways it may undermine forensic conclusions (e.g., Kassin et al., 2013), and the ways it intersects with other biases (e.g., adversarial allegiance; Murrie & Boccaccini, 2015). In fact, two of the eight best practices for valid forensic psychological assessment are to explicitly consider limitations and assumptions and to explicitly weigh alternative views and disagreements (Neal, Martire, et al., 2022). Inconsistent and disconfirming data or results likely limit or qualify an expert's opinion, at minimum. Showing one's work in these ways is a critical component of a method that

enables consumers of expert reports and testimony to understand the scope and nature of the expert's opinion and how to weigh and rely on it while trying the overall facts of the case (e.g., Cunliffe & Edmond, 2021; Edmond, 2015; National Research Council, 2009).

We encourage adopting strategies from cognitive psychology (see prior section) that may appear tedious to evaluators but promote greater gathering and consideration of data. For example, for every component of the legal standards for competence or sanity, list evidence for and against. For every item in a structured professional judgment measure of risk, psychopathic personality, or other forensic-relevant constructs, list the evidence for and against assigning a point(s) on the item (creating a simple two-column table can help evaluators with this process). Indeed, some of this "for and against" information—compiled during the course of evaluation, instrument scoring, opinion formation, and so on—should ultimately appear in the written reports of forensic evaluation, to make clear to readers that evaluators tried to "weigh all . . . rival hypotheses impartially" (American Psychological Association, 2013, p. 3).

Of course, a robust comparison of confirmatory and disconfirmatory information requires that evaluators have conducted a robust investigation to begin with. Prioritizing interview questions that might yield unexpected or disconfirmatory information is crucial. Evaluators should consider which potential records, or collateral interviewees, might be most likely to reveal an alternate, unexpected, or contradictory perspective on the examinee. Overreliance on one source of information—whether interview or a particular portion of the record base—is likely to increase the risk of confirmation bias.

Limitations

Although the results of this project enhance our understanding of psychologists' clinical decision making and potential bias, especially confirmatory bias, the studies are not without

limitations and results should be considered in this context. These studies involve simulation (studies designed to simulate real-world clinical decision making), which has inherent limitations. For example, in actual evaluations, there is no forced choice on availability of information, more information is available than a short summary of a vignette, and there are opportunities to interview the evaluee and speak to collateral sources with relevant information.

As this is an experimental study, it has high internal validity such that causal inferences can be made (which is difficult or impossible to understand from observing real-world clinical work); however, internal validity often comes at the cost of external validity. Indeed, although we designed these studies to simulate clinical decision making, it is impossible to accurately simulate an evaluation with vignettes and limited forced-choice options. Therefore, the real-world applicability of the findings may be limited. The possibility that a clinician planned to ask disconfirmatory information later cannot be ruled out. That is, during the time-limited vignette-style questioning, the clinician in the study may have sought information to narrow down their diagnosis but only as a part of their otherwise comprehensive information-gathering process that may have eventually included questions that would yield disconfirmatory information.

Despite the strong experimental control in this study, it is possible that the paradigm was not complex enough to elicit some of the cognitive biases that may affect clinicians in their real-world practice. Psychologists usually operate in decision environments in which too much information is available to them, and they must sift through all the relevant and irrelevant information for what they need. Order and context effects tend to occur when there is too much information available to remember and the task is complex—here, the vignette was short and the number of diagnostic options was limited. It may be that, in a more ecologically valid decision context without artificial information loss, order and context effects would matter. This

observation also suggests that limiting the information available to just what is relevant and necessary has bias-mitigating potential for psychologists' diagnostic reasoning processes.

Relatedly, pilot testing the experimental study materials could have helped ensure that the manipulations were sufficiently salient. The lack of findings regarding contextual effects could mean that the manipulations were not sufficiently strong enough to detect them, rather than that there are no contextual effects in diagnostic reasoning about which to be concerned. It could have been useful in pilot testing—or in the study, if sample size had allowed it—to have a no-context condition, to determine the diagnostic base rate for the vignette alone without the evaluation context.

There may be constraints on the generalizability of the findings given these particular samples of participants. Our population of interest was licensed psychologists in the United States, with a lesser but relevant interest in psychologists who do forensic work. In the United States, psychologists are licensed as generalists, but they can choose to specialize and work with specific populations or issues. Our sampling procedure better captured the latter than the former: Our sampling database included 2,221 licensed psychologists with primary or secondary forensic interests in the United States. Of note, however, both of our samples reported spending more of their time on clinical nonforensic activities (50% and 49% in Studies 1 and 2, respectively) than forensic activities (30% and 29% in Studies 1 and 2, respectively). Broadly, Lin et al. (2016) estimated that there are 106,000 licensed psychologists in the United States. Of that population, 7% have a primary or secondary specialty in forensic psychology (Lin et al., 2017). Thus, our database captured about 30% of the estimated 7,420 licensed psychologists in the United States with a primary or secondary specialty in forensic psychology and about 2% of the overall population of licensed psychologists in the United States.

With regard to the generalizability of the characteristics of these particular samples to the population of interest, sample demographics can be compared with other studies of the same population, including those using different sampling methods. Across multiple features (i.e., age, years of experience in the field, employment setting, gender, and race), the samples of the studies reported in this article are similar to other studies sampling from the same population (e.g., Neal & Brodsky, 2016; Neal & Grisso, 2014a; Neal & Line, 2022).

Specifically, the average age of participants in these samples (53 years and 59 years in Studies 1 and 2, respectively) is similar to the average age reported in other studies of this same population, including those using different sampling methods (e.g., average age of 59 years in study by Neal & Brodsky, 2016; average age of 53 years in study by Neal & Line, 2022). Years of experience in the field in the current studies (19 years and 21 years on average in Studies 1 and 2, respectively) are similar to the 22 years of average experience reported by Neal and Brodsky (2016), 20 years reported by Neal and Line (2022), and 17 years reported by Neal and Grisso (2014a). Despite the fact that the current samples are similar in these ways across other studies of the same population, it is possible that all of these samples skew toward older psychologists who have been in practice for a longer time.

Most of the participants in both of the current samples worked in private practice (58% and 53% in Studies 1 and 2, respectively), which is similar to some other studies drawing from the same population (e.g., 58% in study by Neal & Brodsky, 2016) but somewhat higher than others (e.g., 48% in study by Neal & Line, 2022). Somewhat fewer psychologists in the current studies worked in institutions or agencies (8% and 13%, respectively) compared with other samples of the same population (e.g., 18% in study by Neal & Brodsky, 2016; 21% in study by Neal & Line, 2022).

With regard to gender, more women than men participated in Study 1 (34% vs. 44%), but fewer women than men participated in Study 2 (24% vs. 53%). Previous studies with samples from this population have reported fewer women than men (e.g., 31% vs. 69% in study by Neal & Brodsky, 2016), and others have reported approximately even participation of women and men (e.g., 45% vs. 42% in study by Neal & Line, 2022). More women than men are currently entering psychological practice in general and forensic psychology in particular, but historically, more men than women practiced in the field at large and in forensic psychology specifically (for further discussion of the shifting gender demographics of the field, see Neal & Line, 2022). Regarding race, the high proportion of White people in these samples is similar to what is observed in other samples from the same population (e.g., Neal & Brodsky, 2016; Neal & Line, 2022) and likely reflects the limited diversity of the population of interest.

Finally, during the creation of the contact database, our research assistants were able to find postal mail addresses and email addresses for most psychologists we identified as relevant for the purposes of our database through the search process of states' licensing databases. But for some, we could find only one type of contact information or the other. Thus, we used two different recruitment methods (email and postal mail) across these two studies to better sample from our full database. The response rates of these studies were similar to those of other studies involving this same population (e.g., 21% response rate from email invitations in study by Neal & Line, 2022) but lower than some earlier studies (e.g., 42% response rate from postal invitations in study by Neal & Brodsky, 2016). The response rates in these studies of 22% (email invitation in Study 1) and 17% (postal mail invitation in Study 2) raise the possibility of response bias, such that the people who opted in were perhaps more interested in a study described as investigating "how practitioners make clinical decisions regarding diagnosis" than people who

did not respond, or perhaps they had more time to respond than people who did not and, thus, could have responded in some systematically different way from people less interested in the topic of diagnostic decision making or who did not have approximately 20 min available to participate.

Conclusion

Bias remains a concern in psychological practice as well as in expert judgments more generally. Unfortunately, psychology experts often fail to implement strategies to manage and reduce bias (Neal & Brodsky, 2016). Results of these studies show small biasing effects from the order of symptom presentation as well as robust evidence of confirmatory information seeking after forming an initial diagnosis. In other words, after psychologists formed a diagnostic hypothesis, they were prone to seek information that supported that hypothesis. Although further research is needed to better understand bias, these findings validate previous works (see Arkes et al., 1988; Lerner & Tetlock, 1999; Mussweiler et al., 2000) that suggest psychologists need to take specific steps to reduce bias in their work and that the field needs to develop standardized, evidence-based processes to mitigate these effects.

References

Acklin, M. W., Fuger, K., & Gowensmith, W. (2015). Examiner agreement and judicial consensus in forensic mental health evaluations. *Journal of Forensic Psychology Practice*, *15*(4), 318–343. https://doi.org/10.1080/15228932.2015.1051447

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.).

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.).

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorder*s (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, *68*(1), 7–19. https://doi.org/10.1037/a0029889

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. https://doi.org/10.1037/0021-9010.73.2.305

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*(3), 258–290. https://doi.org/10.1037/h0055756

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*, 43–63. https://doi.org/10.1002/jip.19

Barefoot v. Estelle, 463 U.S. 880 (1983).

Barrows, H. S., Feightner, J. W., Neufeld, V. R., & Norman, G. R. (1978). *Analysis of the clinical methods of medical students and physicians*. McMaster University School of Medicine.

Brooner, R. K., Schmidt, C. W., Felch, L. J., & Bigelow, G. E. (1992). Antisocial behavior of intravenous drug abusers: Implications for the diagnosis of antisocial personality disorder. *American Journal of Psychiatry*, *149*, 482–487. https://doi.org/10.1176/ajp.149.4.482

Cassidy, M. F., & Buede, D. (2009). Does the accuracy of expert judgment comply with common sense? Caveat emptor. *Management Decision*, *47*(3), 454–469. https://doi.org/10.1108/00251740910946714

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Conway, C. C., Krueger, R. F., & HiTOP Consortium Executive Board. (2021). Rethinking the diagnosis of mental disorders: Data-driven psychological dimensions, not categories, as a framework for mental-health research, treatment, and training. *Current Directions in Psychological Science*, *30*(2), 151–158. https://doi.org/10.1177/096372142199035

Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors*, *50*(5), 745–754. https://doi.org/10.1518/001872008X354183

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, *297*, 35–46. https://doi.org/10.1016/j.forsciint.2019.01.016

Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review*, *116*(4), 971–985. https://doi.org/10.1037/a0017050

Cunliffe, E., & Edmond, G. (2021). Justice without science? Judging the reliability of forensic science in Canada. *Canadian Bar Review*, *99*, 65–112. https://heinonline.org/HOL/P?h=hein.journals/canbarev99&i=65

Cunningham, M. D., & Reidy, T. J. (1998). Antisocial personality disorder and psychopathy: Diagnostic dilemmas in classifying patterns of antisocial behavior in sentencing evaluations. *Behavioral Sciences & the Law*, *16*(3), 333–351. https://doi.org/10.1002/(SICI)1099-0798(199822)

Drew, T., Võ, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*(9), 1848–1853. https://doi.org/10.1177/0956797613479386

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, *51*(4), 204–208. https://doi.org/10.1016/j.scijus.2011.08.004

Edmond, G. (2015). Forensic science evidence and the conditions for rational (jury) evaluation. *Melbourne University Law Review*, *39*(1), 77–127. https://doi.org/10.3316/informit.504529093174012

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press.

Estelle v. Smith, 451 U.S. 454 (1981).

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*, 239–260. https://doi.org/10.1037/0033-295X.90.3.239

Forbes, M. K., Neo, B., Nezami, O. M., Fried, E. I., Faure, K., Michelsen, B., Twose, M., & Dras, M. (2024). Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychological Medicine*, *54*(5), 886–894. https://doi.org/10.1017/S0033291723002544

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*(3), 522–559. https://doi.org/10.1037/pspa0000160

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, *52*, 467–491. https://doi.org/10.1086/269125

Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, *8*, 109–130. https://doi.org/10.1146/annurev-clinpsy-032511-143102

Fraser-Mackenzie, P. A., Dror, I. E., & Wertheim, K. (2013). Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments. *Science & Justice*, *53*(2), 144–153. https://doi.org/10.1016/j.scijus.2012.12.002

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Growns, B., & Neal, T. M. S. (2024). Forensic science decision-making: Expertise lends both skills and vulnerabilities. In M. K. Miller, L. A. Yelderman, J. A. Cantone, & M. T. Huss (Eds.), *The Cambridge handbook of psychology and legal decision-making* (pp. 193–210). Cambridge University Press. https://osf.io/preprints/psyarxiv/m8j4z

Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford University Press.

Heilbrun, K., & Brooks, S. (2010). Forensic psychology and forensic science: A proposed agenda for the next decade. *Psychology, Public Policy, and Law*, *16*(3), 219–253. https://doi.org/10.1037/a0019138

Hoffrage, U., & Gigerenzer, G. (2004). How to improve the diagnostic inferences of medical experts. In E. Kursz-Milcke & G. Gigerenzer (Eds.), *Experts in science and society* (pp. 249–268). Kluwer Academic.

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

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557–571. https://doi.org/10.1037/0022-3514.80.4.557

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, *14*(2), 147–174. https://doi.org/10.1080/13669877.2010.511246

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus & Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

King, K. A., & Vaughan, J. L. (2004). Influence of paper color and a monetary incentive on response rate. *Psychological Reports*, *95*, 432–434.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Krane, D. E., Ford, S., Gilder, J. R., Inman, K., Jamieson, A., Koppl, R., Kornfield, I. L., Risinger, M. D., Rudin, N., & Taylor, M. S. (2008). Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Sciences*, *53*(4), 1006–1007. https://doi.org/10.1111/j.1556-4029.2008.00787.x

Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*(3), 242–247. https://doi.org/10.1177/1745691613483477

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0023709

Langenburg, G., Champod, C., & Wertheim, P. (2009). Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences*, *54*(3), 571–582. https://doi.org/10.1111/j.1556-4029.2009.01025.x

LeBlanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: The influence of diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine: Journal of the Association of American Medical Colleges*, *77*(Suppl.), S67–S69. https://doi.org/10.1097/00001888-200210001-00019

Lejarraga, T., & Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychological Bulletin*, *147*(6), 535–564. https://doi.org/10.1037/bul0000324

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Lilienfeld, S. O., & Treadway, M. T. (2016). Clashing diagnostic approaches: DSM-ICD versus RDoC. *Annual Review of Clinical Psychology*, *12*, 435–463. https://doi.org/10.1146/annurev-clinpsy-021815-093122

Lin, L., Christidis, P., & Stamm, K. (2016, March). Datapoint: Where are the highest concentrations of licensed psychologists?: News from APA's Center for Workforce Studies. *American Psychological Association Monitor on Psychology*, *47*, 13. https://www.apa.org/monitor/2016/03/datapoint

Lin, L., Christidis, P., & Stamm, K. (2017, September). Datapoint: A look at psychologists' specialty areas: News from APA's Center for Workforce Studies. *American Psychological Association Monitor on Psychology*, *48*, 15. https://www.apa.org/monitor/2017/09/datapoint

Ludolph, R., & Schulz, P. J. (2018). Debiasing health-related judgments and decision making: A systematic review. *Medical Decision Making*, *38*(1), 3–13. https://doi.org/10.1177/0272989X17716672

MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287. https://doi.org/10.1146/annurev.psych.49.1.259

MacLean, N., Neal, T., Morgan, R. D., & Murrie, D. C. (2019). Forensic clinicians' understanding of bias. *Psychology, Public Policy, and Law*, *25*(4), 323–330. https://doi.org/10.1037/law0000212

Marsh, J. K., De Los Reyes, A., & Lilienfeld, S. O. (2018). Leveraging the multiple lenses of psychological science to inform clinical decision making: Introduction to the special section. *Clinical Psychological Science*, *6*(2), 167–176. https://doi.org/10.1177/2167702617736853

Mattijssen, E. J. A. T., Witteman, C. L. M., Berger, C. E. H., & Stoel, R. D. (2020). Cognitive biases in the peer review of bullet and cartridge case comparison casework: A field study. *Science & Justice*, *60*(4), 337–346. https://doi.org/10.1016/j.scijus.2020.01.005

Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2017). *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*. Guilford.

Miller, L. S. (1984). Bias among forensic document examiners: A need for procedural change. *Journal of Police Science & Administration*, *12*(4), 407–411.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights From the Behavioral and Brain Sciences*, *2*(1), 129–140. https://doi.org/10.1177/2372732215600886

Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science*, *11*, 37–55. https://doi.org/10.1146/annurev-lawsocsci-120814-121714

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, *24*(10), 1889–1897. https://doi.org/10.1177/0956797613481812

Mussweiler, T., Strack, F., & Pfeif, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142–1150. https://doi.org/10.1177/01461672002611010

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. National Academies Press. https://doi.org/10.17226/12589

Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *22*(1), 58–76. https://doi.org/10.1037/law0000077

Neal, T. M. S., & Grisso, T. (2014a). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice & Behavior*, *41*(12), 1406–1421. https://doi.org/10.1177/0093854814548449

Neal, T. M. S., & Grisso, T. (2014b). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *20*(2), 200–211. https://doi.org/10.1037/a0035824

Neal, T. M. S., Lienert, P., Denne, E., & Singh, J. P. (2022). A general model of cognitive bias in human judgment and systematic review specific to forensic mental health. *Law and Human Behavior*, *46*(2), 99–120. https://doi.org/10.1037/lhb0000482

Neal, T. M. S., & Line, E. N. (2022). Income, demographics, and life experiences of clinical-forensic psychologists in the United States. *Frontiers in Psychology*, *13*, Article 910672. https://doi.org/10.3389/fpsyg.2022.910672

Neal, T. M. S., Martire, K. A., Johan, J. L., Mathers, E., & Otto, R. K. (2022). The law meets psychological expertise: Eight best practices to improve forensic psychological assessments. *Annual Review of Law and Social Science*, *18*, 169–192. https://doi.org/10.1146/annurev-lawsocsci-050420-010148

Neal, T. M. S., & Pronin, E. (2023, September 29). *Project outcomes report: Measuring the objectivity and bias of experts* (U.S. National Science Foundation Award No. 1655011). https://www.nsf.gov/awardsearch/showAward?AWD_ID=1655011

Neal, T. M. S., Slobogin, C., Saks, M. J., Faigman, D., & Geisinger, K. (2019). Psychological assessments in legal contexts: Are courts keeping "junk science" out of the courtroom? *Psychological Science in the Public Interest*, *20*(3), 135–164. https://doi.org/10.1177/1529100619888860

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Osborne, N. K., Taylor, M. C., Healey, M., & Zajac, R. (2016). Bloodstain pattern classification: Accuracy, effect of contextual information and the role of analyst characteristics. *Science & Justice*, *56*(2), 123–128. https://doi.org/10.1016/j.scijus.2015.12.005

Osborne, N. K., Woods, S., Kieser, J., & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science & Justice*, *54*(4), 267–273. https://doi.org/10.1016/j.scijus.2013.12.005

Popper, K. R. (1959). The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, *10*(37), 25–42.

President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

Pronin, E., & Hazel, L. (2023). Humans' bias blind spot and its societal significance. *Current Directions in Psychological Science*, *32*(5), 402–409. https://doi.org/10.1177/09637214231178745

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*(4), 565–578. https://doi.org/10.1016/j.jesp.2006.05.011

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Quigley-McBride, A., Dror, I. E., Roy, T., Garrett, B. L., & Kukucka, J. (2022). A practical tool for information management in forensic decisions: Using Linear Sequential Unmasking-Expanded (LSU-E) in casework. *Forensic Science International: Synergy*, *4*, Article 100216. https://doi.org/10.1016/j.fsisyn.2022.100216

Regier, D. A., Farmer, M. E., Rae, D. S., Locke, B. Z., Keith, S. J., Judd, L. J., & Goodwin, F. K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse: Results from the Epidemiologic Catchment Area (ECA) study. *Journal of the American Medical Association*, *264*, 2511–2518. https://doi.org/10.1001/jama.1990.03450190043026

Rogers, R., Duncan, J. C., Lynett, E., & Sewell, K. W. (1994). Prototypical analysis of antisocial personality disorder. *Law and Human Behavior*, *18*, 471–484. https://doi.org/10.1007/BF01441386

Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: A systematic review. *BMC Medical Informatics and Decision Making*, *16*(1), Article 138. https://doi.org/10.1186/s12911-016-0377-1

Schulz-Hardt, S., Frey, D., Luthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality & Social Psychology*, *78*, 655–669. https://doi.org/10.1037/0022-3514.78.4.655

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*(2), 252–266. https://doi.org/10.1016/0749-5978(92)90064-E

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, *63*, 129–138.

Skodol, A. E., Oldham, J. M., & Gallaher, P. E. (1999). Axis II comorbidity of substance use disorders among patients referred for treatment of personality disorders. *American Journal of Psychiatry*, *156*, 733–738. https://doi.org/10.1176/ajp.156.5.733

Sneyd, D., Schreiber Compo, N., Rivard, J., Pena, M., Stoiloff, S., & Hernandez, G. (2020). Quality of laypersons' assessment of forensically relevant stimuli. *Journal of Forensic Sciences*, *65*(5), 1507–1516. https://doi.org/10.1111/1556-4029.14495

Stevenage, S. V., & Bennett, A. (2017). A biased opinion: Demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results. *Forensic Science International*, *276*, 93–106. https://doi.org/10.1016/j.forsciint.2017.04.009

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113. https://doi.org/10.1017/S1930297500007622

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. DOI: 10.1126/science.745568

Varga, A. L., & Hamburger, K. (2014). Beyond type 1 vs. type 2 processing: The tri-dimensional way. *Frontiers in Psychology*, *5*, Article 993. https://doi.org/10.3389/fpsyg.2014.00993

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*(3), 506–519. https://doi.org/10.1037/a0028857

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, *118*(2), 357–378. https://doi.org/10.1037/a0022936

Zapf, P. A., & Dror, I. E. (2017). Understanding and mitigating bias in forensic evaluation: Lessons from forensic science. *International Journal of Forensic Mental Health*, *16*(3), 227–238. https://doi.org/10.1080/14999013.2017.1317302

Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheeles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice*, *4*(1), 27–44. https://doi.org/10.1300/J158v04n01_02

Figure 1. *Neal, Lienert et al., (2022)'s Descriptive Model of Bias in Human Judgment*

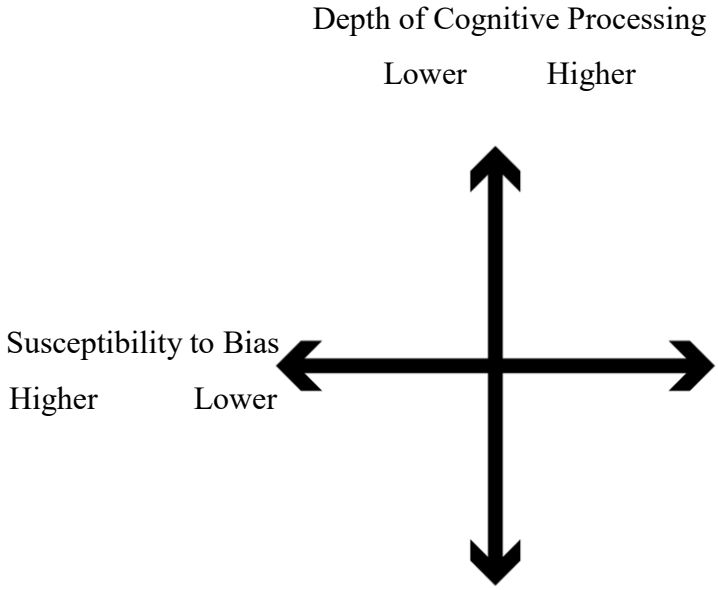
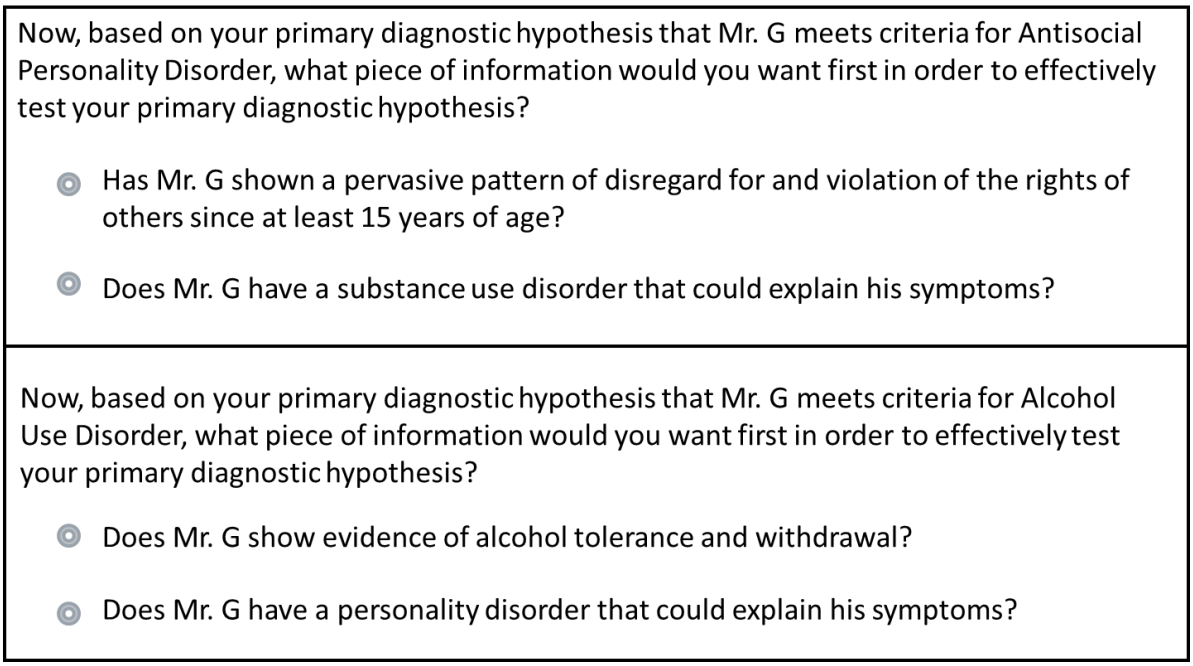
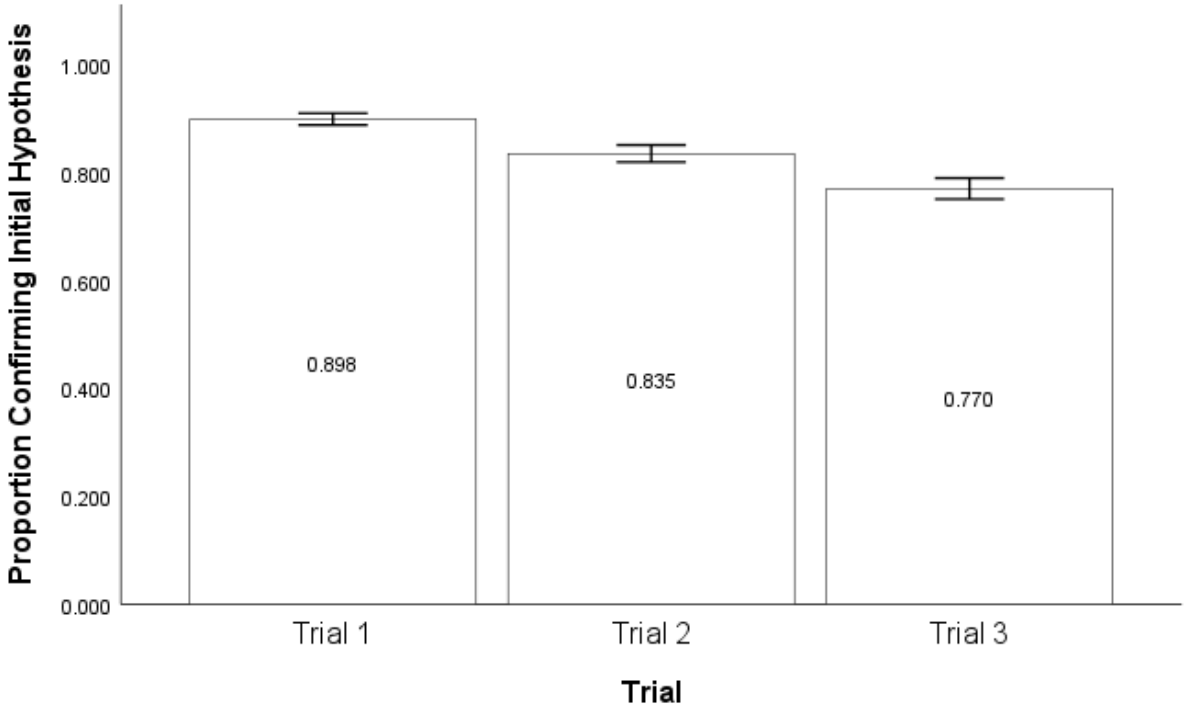


Figure 2. *Two of the Piped Follow-Up Questions Linked to the Diagnostic Hypothesis Participants Rank-Ordered First in Study 1*



Note: After selecting their initial diagnostic hypothesis in Study 1, psychologist-participants received a piped follow-up question based on that initial hypothesis. We designed four piped questions, one for each possible initial diagnostic hypothesis (participants only saw one follow-up question – the one designed specifically for their diagnostic selection). Two (the most vignette-relevant) of the four questions we designed are shown in this figure. Each question offered a choice between two pieces of information: one that might *confirm* the psychologists' initial hypothesis, and one that might *disconfirm* it. The questions shown here present the potentially confirmatory information first and the potentially disconfirmatory information second, but their orders were randomized for participants.

Figure 3. *Proportion of Psychologist-Participants Choosing Confirmatory Information at Each of the Three Sequential Decision Opportunities in Study 2*



Note: Generalized logistic mixed effects model with choice opportunity as predictor (trial) and random intercept for psychologist-participant; error bars represent the 95% confidence intervals for the predicted probability of a confirming choice.

Table 1. *Order and Context Effects in Studies 1 and 2: Proportions of Initial Diagnostic Hypotheses and Sample Sizes (*n*)*

Study	Initial Diagnostic Hypothesis	Order of Symptoms		Context: Referral Source	
		"Short Fuse" 1st	Alcohol 1st	Work	Court
Study 1	Antisocial Personality Disorder	0.38	0.28	0.34	0.32
	Alcohol Use Disorder	0.62	0.72	0.66	0.68
	n	74	68	71	71
Study 2	Antisocial Personality Disorder	0.46	0.26	0.31	0.40
	Alcohol Use Disorder	0.54	0.74	0.69	0.60
	n	59	65	61	63

Table 2. *Study 2 Generalized Mixed Effects Logistic Regression Model of Confirmatory Information Seeking Across Three Decision Opportunities*

	Coefficient	Standard Error	Test Statistic	*p*	Standardized Effect Size
Fixed Effects					Odds Ratio [95% CI]
Intercept	2.31 [1.69, 2.93]	0.32	$t = 7.34$	<0.001	10.11 [5.44, 18.78]
Trial 2	-0.59 [-1.76, -0.30]	0.39	$t = -1.52$	0.129	0.56 [0.26, 1.19]
Trial 3	-1.03 [-1.35, 0.17]	0.37	$t = -2.76$	0.006	0.36 [0.17, 0.74]
Random Effects					Cohen's *d*
Participants (*N*=128)	0.54 [0.12, 2.39]	0.41	$Z = 1.32$	0.186	0.12 (very small)

Note: Generalized mixed effects logistic regression model with choice opportunity as predictor (three trials) and random intercept for psychologist-participant on probability of seeking confirming information vs. disconfirming information. Trial 1 was the reference category. Values in brackets are 95% confidence intervals.

Table 3. *Study 2 Correlation Matrix*

Measure	BBS	BBS_Frame *r* *(p)*	BBS_Context *r* *(p)*	BBS_Confirm *r* *(p)*	#Confirm *r* *(p)*	CRT-3 *r* *(p)*	CRT-11 *r* *(p)*
BBS	-	.23 (.015)	-.06 (.53)	.06 (.52)	.10 (.29)	-.01 (.93)	-.15 (.13)
BBS_Frame		-	.19 (.045)	.20 (.031)	.02 (.82)	-.21 (.03)	-.19 (.06)
BBS_Context			-	.11 (.25)	-.02 (.87)	-.09 (.36)	-.21 (.040)
BBS_Confirm				-	-.08 (.42)	-.04 (.71)	.04 (.73)
#Confirm					-	.06 (.56)	.01 (.89)
CRT-3						-	.75 (<.001)
CRT-11							-

Note. BBS (Bias Blind Spot, operationalized as rating of others' bias – ones' own bias). We operationalized the other BBS items (Frame, Context, Confirm) similarly but specific to each type of bias. #Confirm (total number of confirming choices made Study 2). CRT (Cognitive Reflection Task, 3-item scale and 11-item scale).