Understanding and Mitigating Bias in Forensic Evaluation: Lessons from Forensic Science

Patricia A. Zapf and Itiel E. Dror

ABSTRACT
Criticism has emerged in the last decade surrounding cognitive bias in forensic examinations. The National Research Council (NRC, 2009) issued a report that delineated weaknesses within various forensic science domains. The purpose of this article is to examine and consider the various influences that can bias observations and inferences in forensic evaluation and to apply what we know from forensic science to propose possible solutions to these problems. We use Sir Francis Bacon’s doctrine of idols—which underpins modern scientific method—to expand Dror’s (2015) five-level taxonomy of the various stages at which bias can originate within forensic science to create a seven-level taxonomy. We describe the ways in which biases can arise and impact work in forensic evaluation at these seven levels, highlighting potential solutions and various means of mitigating the impact of these biases, and conclude with a proposal for using scientific principles to improve forensic evaluation.

KEYWORDS
Bias; cognitive bias; cognitive factors; forensic evaluation; forensic psychology

Since the NRC Report, several researchers and commentators have written about the impact of cognitive biases in the forensic sciences (e.g., Dror, 2015) and have proposed solutions to mitigate the impact of these biases (Dror, Thompson, Meissner, Kornfield, Krane, Saks & Risinger, 2015; Found & Ganas, 2013; Krane et al., 2008, Mattijssen et al., 2015). The overarching goal in all of this, of course, is to make the various disciplines within the field of forensics as scientific as possible and to limit the impact of factors that bias examiners and might impede impartial and objective observations and conclusions.

While we acknowledge differences between the workflow and roles of various forensic science practitioners and forensic mental health evaluators, we also believe that there are overarching similarities in the tasks required between the forensic science and forensic mental health evaluation domains. Across these two domains, examiners and evaluators are tasked with collecting and considering various relevant pieces of data in arriving at a conclusion or opinion and, across both of these domains, irrelevant information can change the way an examiner/evaluator interprets the relevant data. Bias mechanism, such as bias cascade and bias snowball, can impact examiners in forensic science as well as in
forensic psychology (Dror, Morgan, Rando, & Nakhaei-zadeh, in press). We also acknowledge that completely unbiased and impartial examinations (forensic science) and evaluations (forensic mental health) are aspirational in nature and that, because examiners/evaluators are human, it is impossible to eliminate all bias. Nevertheless, the intent here is to shed light on the issue of bias, the various stages at which it can exert influence, and possible ways in which to think about minimizing its impact on decisions and outcomes in forensic evaluation.

The purpose of this article is to examine and consider the various influences that can bias observations and inferences in forensic evaluation and to apply what we know from forensic science to propose possible solutions to these problems. We use Sir Francis Bacon’s doctrine of idols—which underpins modern scientific method—to expand a five-level taxonomy (Dror, 2015) of the various stages at which bias can originate within forensic science to create a seven-level taxonomy applicable to forensic evaluation. We describe the ways in which biases can arise and impact work in forensic evaluation at these various levels, highlighting potential solutions and various means of attempting to mitigate the impact of these biases, and conclude with a proposal for next steps on the path forward with the hope that increased awareness of and exposure to these issues will continue to stimulate further research and discussion in this area.

Sir Francis Bacon, who laid the foundations for modern science, believed that scientific knowledge could only arise if we avoid factors that distort and prevent objectivity. Nearly 400 years ago, Bacon developed the doctrine of “idols,” in which he set out the various obstacles that he believed stood in the way of truth and science—false idols that prevent us from making accurate observations and achieving understanding by distorting the truth and, therefore, stand in the way of science. These obstacles were categorized into idola tribus (idols of the tribe), idola spectus (idols of the den or cave), idola fori (idols of the market), and idola theatrae (idols of the theater). The implications of these obstacles for forensic science have been described in Dror (2009). For example, Bacon makes the case that experiences, education, training, and other personal traits (the idola spectus) that derive from nurture, can cause people to misperceive and misinterpret nature differently. That is, because of individual differences in their upbringing, experiences, and professional affiliations, people develop personal allegiances, ideologies, theories, and beliefs, and these may “corrupt the light of nature” (Bacon, 1620/1902, p. 21).

Bacon’s doctrine of idols distinguishes between idols that are a result of our physical nature (e.g., human cognitive architecture) and the ways in which we were nurtured (e.g., experiences), and those that result from our social nature and the fact that we are social animals who interact with others in communities and work together. The first two idols—those of the tribe and the den—result from our physical nature and upbringing respectively, whereas the others—those of the market and theater—result from our social nature and our interactions with others. Although we acknowledge an imperfect alignment between the idols as Bacon conceptualized them and how they might be conceptualized and interpreted today, and indeed including how we might have done so in this article, this conceptualization nonetheless provides a useful framework within which to consider challenges and obstacles in various domains. Here we elaborate on Dror’s (2009) use of this framework for forensic science by integrating it into a discussion of the challenges and obstacles facing forensic evaluation.

In parallel and in addition to Bacon’s four idols, Dror and his colleagues have discussed various levels at which cognitive factors might interfere with objective observations and inferences and contribute to bias within the forensic sciences (see Dror, 2015; Stoel, Berger, Kerkhoff, Mattijsen, & Dror, 2015). These various levels encompass both case-specific components as well as more general, base-rate, and organizational factors. When considering Bacon’s doctrine of idols in light of the various factors that might introduce bias into forensic decision-making, it appears that Bacon’s idols can be integrated into these levels to further expand and enrich the taxonomy. Here we present a seven-level taxonomy that integrates Bacon’s doctrine of idols with the previous work of Dror and colleagues on the various sources of bias that might be introduced, and apply these to forensic evaluation.

**Forensic evaluation as akin to scientific investigation**

Forensic evaluation requires the collection and examination of various pieces of data to arrive at an opinion regarding a particular legal issue at hand. The most common type of forensic evaluation conducted is the examination of a defendant’s competency to stand trial, with more than 60,000 evaluations of this type being ordered each year in the United States (see Melton, Petrila, Poythress, & Slobozyn, 2007) and relatively similar proportions (taking population statistics into account) being ordered in countries with adversarial systems of justice, such as Canada, the United Kingdom, and Australia (Reed & Zapf, in press). Other types of forensic evaluations include assessments of a defendant’s mental state at the time of the offense, assessments of risk for general violence or specific forms of violence (e.g., intimate
partner violence, sexual violence), and assessments of other issues relevant to either criminal or civil law (e.g., evaluation of competence to waive Miranda rights, evaluation of parenting capacity, evaluation of testamentary capacity, evaluation of capacity to provide informed consent to treatment).

The common components of all forensic evaluations include the collection of data relevant to the issue at hand—typically including official (police) reports and third-party or collateral information sources relevant to the legal issue being evaluated, interview data provided by the individual being evaluated, and the results of psychological testing, if relevant to the legal inquiry—and the consideration and weighting of these various pieces of data, according to relevance and information source, to arrive at an opinion/conclusion regarding the legal issue being evaluated (see Neal & Grisso, 2014a for a survey of the types of tools and data used in various types of forensic evaluations). Forensic evaluation is distinct from clinical evaluation, which relies primarily on limited self-report data from the individual being evaluated. Forensic evaluation places great importance on collecting and considering third party and collateral information in conjunction with an evaluee’s self-report data, and forensic evaluators are expected to consider the impact and relevance of the various pieces of data on their overall conclusions (see Melton et al., 2007). In addition, forensic evaluators are expected to strive to be as impartial, objective, and unbiased as possible in arriving at their conclusions and opinions about the legal issue at hand. Indeed, forensic evaluators are trained to write reports that delineate the data that were considered, the inferences made on the basis of those data, and the opinions or conclusions arrived at on the basis of the data and the inferences made from those data (see, for example, Grisso, 2010; Otto, DeMier, & Boccaccini, 2014). Hence, it can be argued that forensic evaluations should aspire to be more similar to scientific investigations—where the emphasis is placed on using observations and data to test alternate hypotheses—than to unstructured clinical assessments, which accept an evaluee’s self-report at face value without attempts to corroborate or confirm the details of the evaluee’s account and with less emphasis on alternate hypothesis testing (additional details on the differences between forensic evaluation and clinical evaluation can be found in Goldstein (2003) and in Melton, Petrila, Poythress, & Slobogin (2007)).

If we accept the premise that forensic evaluations should be more akin to scientific investigations than clinical evaluations, then forensic evaluators should conduct their work more like scientists than clinicians, using scientific methods to inform their conceptualization of the case and opinions regarding the legal issue at hand. Maximizing the accuracy and objectivity of observations and inferences is crucial to arriving at minimally biased conclusions regarding the legal issue being evaluated. We take the lessons from forensic science and apply these to forensic evaluation with the aim of making forensic evaluation as objective and scientific as possible within the confines and limitations of attempting to apply group-to-individual inferences (see Faigman, Monahan, & Slobogin, 2014 for a detailed discussion of group-to-individual inference in scientific expert testimony). We do so by developing the framework of a seven-level taxonomy delineating the various influences that might interfere with objective observations and inferences, potentially resulting in biased conclusions in forensic evaluation. The taxonomy starts at the bottom with innate sources that have to do with being human. As we ascend the taxonomy, we discuss sources related to nurture—such as experience, training, and ideology—that can cause bias and, as we near the top of the taxonomy, the sources related to the specific case at hand. So, the order of the taxonomy is from general, basic innate sources derived from human nature, to sources that derive from nurture, and then to those sources that derive from the specifics of the case at hand (see Figure 1).

**Influences that might interfere with objective observations and inferences**

Factors that might interfere with objective observations and inferences, and thus lead to potentially biased decision making, exist at a number of levels and for various reasons. Figure 1 delineates the seven-level taxonomy that summarizes the various influences that might interfere with objective decision-making in forensic evaluation. We conceptualize this taxonomy as spanning from general influences that result from our basic human nature (bottom of the pyramid) to case-specific influences (top of the pyramid), denoted by the bi-directional arrow at the left side of the figure. We discuss each of the levels of the taxonomy within the context of forensic evaluation, drawing from forensic science domains.

**Cognitive architecture and the brain**

At the very base of the taxonomy are potentially biasing influences that result from our basic human nature and the cognitive architecture of the brain. These influences are akin to what Bacon considered *idola tribus*, idols of the tribe; that is, obstacles that derive from the very fact of our being members of the human species. These obstacles or influences result from the way in which our brains are built; “they result from our inability to look at the world from ‘outside’ of ourselves, to see things as
they actually are, instead of seeing them as distorted by our own mental processes” (Dror, 2009, p. 95).

The human brain has a limited capacity to represent and process all of the information presented to it and so it relies upon techniques such as chunking information (binding individual pieces of information into a meaningful whole), selective attention (attending to specific pieces of information while ignoring other information), and top-down processing (conceptually driven processing that uses context to make sense of information) to efficiently process information (Lindsay & Norman, 1977). It is naïve to believe that we perceive the world objectively or that we use bottom-up processing (encoding and interpreting the nature of a stimulus based solely upon the properties of the object) to understand information. Indeed, many studies have demonstrated the power and pervasiveness of top-down, conceptually driven processing whereby we unconsciously weave our knowledge of the world into our processing of it (Nickerson, 1998; Nisbett & Wilson, 1977; Kassin, Dror & Kukucka, 2013). We actively process information by selectively attending to that which we assume to be relevant and interpret this information in light of that which we already know. This top-down processing allows us to efficiently process loads of information and occurs in an automatic manner that is largely outside of our conscious awareness.

Ironically, this automaticity and efficiency—which serves as the bedrock for expertise—also introduces degradation in performance. These assets and vulnerabilities are inherent to the cognitive architecture and constitute two inevitable sides of the same coin” (Dror, 2009, p. 96; see also Dror, 2011).

The cognitive system is such that it processes information in an efficient and effective way; however, the shortcut mechanisms used to do so result in performance degradation. Ample research evidence demonstrates the presence of various types of biases that result from our cognitive architecture (Festinger, 1957; Simon, 1956; Tversky & Kahneman, 1973, 1974). For example, information that we encounter first is more influential than information we encounter later. This anchoring bias (see Tversky & Kahneman, 1974) can result in a forensic evaluator being overly influenced by or giving greater weight to information that is initially presented or reviewed. Thus, initial information communicated to the forensic evaluator by the referring party will likely be more influential, and serve as an anchor for, subsequent information reviewed by the evaluator.

We also have a tendency to overestimate the probability of an event or occurrence when other instances of that event or occurrence are easily recalled. This availability bias (see Tversky & Kahneman, 1973) can result in a forensic evaluator overestimating the likelihood of a particular outcome on the basis of being able to readily recall similar instances of that same outcome. Confirmation bias (see Nickerson, 1998) results from our natural inclination to rush to conclusions that confirm what we want, believe, or accept to be true. Indeed, research in the fingerprint analysis domain has demonstrated this confirmation bias, indicating that latent fingerprint examiners are influenced by contextually irrelevant information when making fingerprint comparisons.
In the forensic evaluation domain, Neal and Grissi (2014b) warn that the confirmation bias can exert its influence on evaluators who share a preliminary opinion before the evaluation is complete by committing the evaluator in a way that makes it difficult to resist or overcome this bias in the final interpretation of the data.

These and other biases that result from our cognitive architecture and the way in which our brains are built are within all of us and are simply a part of our human nature. What is important is that we recognize our limits and cognitive imperfections so that we might try to address them by using countermeasures.

Training and motivation

Moving up the taxonomy, the next three sources of influences that can affect our perception and decision-making result from our environment, culture, and experience. First among these are those influences that are brought about by our upbringing—our training and motivations (e.g., Kunda, 1990). Bacon referred to these obstacles as *idola spectus*, idols of the den or cave, and conceptualized these influences as a function of one’s experiences, education, training, and other personal traits or characteristics. As distinguished from the idols of the tribe or those that result from our human nature, the idols of the den or cave are a result of our nurture. “Different people, based on their specific upbringing, life experiences and professional affiliations, have developed personal allegiances to groups, ideologies, disciplines, theories, or methodologies” (Dror, 2009, p. 99). Our personal motivations and preferences, developed through our upbringing, affect our perception, reasoning, and decision-making (Balcetis & Dunning, 2006).

Motivation comes in many forms and motivational bias has been well documented as a source of erroneous decision making within the context of criminal investigators’ judgments of witness reliability (Ask & Granhag, 2007) and need for cognitive closure (Ask & Granhag, 2005). Just as the motivations of criminal investigators can lead to distortions in the way that they perceive and interpret information, so too can the motivations of forensic evaluators distort or cause bias in their perception and interpretation of relevant (and irrelevant) case information.

Closely related to an individual’s motivations are how one sees oneself and with whom that individual identifies. One particularly salient and concerning influence in this realm for forensic evaluators is that of adversarial allegiance; that is, the tendency to arrive at an opinion or conclusion that is consistent with the side that retained the evaluator. Murrie, Boccaccini, and colleagues have demonstrated that forensic evaluators working for the prosecution assign higher psychopathy scores to the same individual as compared to forensic evaluators working for the defense (Murrie, Boccaccini, Johnson, & Janke, 2008). Similarly, forensic evaluators assign higher scores on actuarial risk assessment instruments—known to be less subjective than other types of risk assessment instruments—when retained by the prosecution and lower scores when retained by the defense (Murrie et al., 2009). This was the case in both field studies as well as in a well-controlled experimental manipulation (Murrie, Boccaccini, Guarnera, & Rufino, 2013). In addition, these researchers also found that adversarial allegiance appears to influence norm selection and reporting practices such that defense-retained evaluators were more likely to endorse reporting practices that conveyed the lowest possible level of risk whereas prosecution-retained evaluators were more likely to endorse practices suggesting the highest possible level of risk (Chevalier, Boccaccini, Murrie, & Varela, 2015).

In addition to the pull to affiliate with the side that retained the forensic evaluator is the issue of pre-existing attitudes that forensic evaluators hold and how these might impact the forensic evaluation process. Neal (2016) surveyed forensic evaluators to determine the impact of attitudes towards capital punishment on willingness to accept referrals from various adversarial parties. She found that evaluators who opposed the death penalty were more likely to decline referrals from any party or accept referrals from the defense only whereas those who supported the death penalty were more likely to accept referrals from any party. This raises the issue of a self-selection bias, as a result of preexisting attitudes, which might contribute to partisan participation in capital evaluations. These issues of pre-existing attitudes, affiliations, and the potential biases they create also relate to the next set of idols described by Bacon, which consider the impact of the communications and interconnections between people.

Organizational factors

The third set of idols described by Bacon—*idola fori*, idols of the market—are a result of our social interactions, communications, and connections with others. These include organizational and cultural factors that reflect the ways in which we understand and communicate with each other, both within and outside of the organization. The primary focus of Bacon’s idols of the market is on the language we use to convey knowledge. Language has a profound effect on how we perceive and think about information. The words we use to convey
knowledge— terminology, vocabulary, and even jargon— can cause errors in how we understand and interpret information when we use them without attention and proper focus on the true meaning, or without definition, measurable criteria, and quantification. It is important to consider the meaning and interpretation of the words we use and how these might differ by organization, discipline, or culture. It is easy to assume that we know what someone means when they tell us something—whether it be an evaluator, a retaining party, or a collateral informant— but we must be cautious about both interpreting the language of others and using language to convey what we mean.

Language and Bacon’s idols of the market also relate to how systems communicate as well as to organizational structures and the flow of information within and between different systems, organizations, or entities. Different methods or procedures can result in different rates of accuracy. For example, differences in the eyewitness lineup identification process have resulted in different rates of accuracy in the conclusions reached (Wells, Steblay, & Dysart, 2015). In the forensic assessment domain, different methods of conducting risk assessments (using dynamic risk assessment measures versus static risk assessment measures) have been demonstrated to affect the predictive accuracy of the conclusions reached by evaluators (Chu, Thomas, Ogloff, & Daffer, 2011). Indeed, several meta-analyses comparing the predictive accuracy of clinical versus actuarial methods for risk assessment have shown that actuarial methods—that is, highly structured methods with explicit decision rules and little room for discretion—outperform unstructured clinical methods and show higher rates of reliability and less bias in the predicted outcomes (e.g., Aegisdottir et al., 2006; Dawes, Faust, & Meehl, 1989; Hanson & Morton-Bourgon, 2009). Structured methods improve reliability and reduce bias by limiting discretion on the part of evaluators; thus, methods allowing or requiring more clinical discretion are more susceptible to evaluator bias. Murrie and colleagues’ (2013) research demonstrated this by finding that an actuarial tool allowing for little discretion (STATIC-99R) demonstrated less bias than a structured tool that preserved some discretion on the part of evaluators/raters (PCL-R).

Combating and safeguarding against these idols of the market should take place at an organizational level, and some institutional structures are better for scientific inquiry than others (see Koppl, 2005). For example, the NRC (2009) report recommended moving forensic laboratories out of police departments. Within existing organizational structures, using language with specific definition and meaning that serves to increase error detection and prevention is important for creating a more scientific discipline. Hawaii’s three-panel system for evaluating competency and insanity of criminal defendants, while not perfect, is one example of an organizational structure that serves to increase error detection and prevention. As Acklin, Fuger, and Gowsensmith (2015) concluded after evaluating examiner agreement and judicial consensus within Hawaii’s three-panel system, procedural standardization, the use of forensic assessment instruments, and the application of structured professional methods can serve to reduce bias in assessments and improve the quality of forensic mental health opinions.

**Base rate expectations**

The final set of idols described by Bacon—*idola theatris*, idols of the theater—are not inherent to our human nature but, rather, result from how we produce knowledge or draw inferences and conclusions. Bacon divided these idols into three types: “*sophistical*, based on just a few anecdotal observations (or even no experimental evidence); *empirical*, based on narrow research; and *superstitious*, based on unsupported or blind belief” (Dror, 2009, p. 106). In considering forensic evaluation in light of these idols of the theater, we must ask what forensic evaluation is based on. That is, to what extent is forensic evaluation based on broad, systematic scientific research and to what extent is it based on anecdotal or narrow, in-house research and observations? To what extent do forensic evaluators accept the science without proper and sufficient questioning? How do they distinguish between what they actually know and what they merely believe? How much of what they do is a matter of belief and how much is a matter of scientific knowledge? These idols are perhaps the most difficult to confront, and defensiveness or a lack of openness to discussing and examining the bases of our knowledge can be serious obstacles that prevent forensic evaluators from broadening and strengthening the scientific basis for their discipline.

The ways in which forensic evaluators produce knowledge within their discipline can serve as an impediment to accurate observations and objective inferences. Anecdotal observations or information based on unsupported or blind beliefs can serve to create expectations about conclusions or outcomes before an evaluation is even conducted. Similarly, using methods or procedures that have not been adequately validated or that have been based on narrow, in-house research for which generalizability is unknown can result in inaccurate conclusions. Drawing inferences on the basis of untested assumptions or base rate expectations can lead to erroneous outcomes.
Early research on the risk and treatability of psychopaths that was narrowly focused on interventions such as milieu therapy with groups of serious, chronic offenders led some to conclude that psychopaths were untreatable (see Douglas, Nikolova, Kelley, & Edens, 2015). As research in this domain expanded to include degrees of psychopathy and various types of interventions, the conclusion that psychopaths are untreatable is no longer widely accepted (see, for example, Salekin 2002). Consideration of that early body of narrowly-focused research sets an expectation for evaluators that individuals displaying psychopathic traits are a high risk to re-offend and have low treatment potential, a different expectation regarding risk and treatment potential than that set by the broader, more diverse body of research that has since accumulated. In either case, creating expectations that might result in potentially biased assessment outcomes for a particular individual.

As we move up the pyramid we move from environmental, cultural, and experiential influences to those influences that are specific to the particular case at hand. That is, the top three levels of the pyramid describe influences that might interfere with accurate observations and objective inferences within a specific forensic evaluation.

Irrelevant case information

Whereas the previous level deals with influences that come about as a result of the way in which we produce knowledge or draw inferences and conclusions from information that does not pertain specifically to the case at hand, this level deals with influences that result from information that is obtained or reviewed for a specific case but that is irrelevant to the referral question. Perhaps one of the most potentially biasing considerations at this level involves the inferences made by others. In any forensic evaluation, the evaluator is tasked with collecting, reviewing, and considering various sources of information and data, much of which is relevant to the referral question but a great deal of which is irrelevant. Detailed information about an evaluee’s criminal history (offenses committed prior to the index offense), in most instances, is irrelevant to the issue of his or her criminal responsibility, which is an inquiry that focuses on the mental state of the individual at the time of the index offense. This irrelevant information, however, can become biasing for an evaluator. Even more potentially biasing can be the inferences and conclusions that others make about an evaluee—including collateral informants as well as retaining and opposing parties—since evaluators typically do not have access to the data or the logic used by others in arriving at these inferences and conclusions.

It is naïve to think that a forensic evaluator can only collect and consider relevant information, especially since many times it is not clear what is relevant and what is irrelevant until all collected materials have been reviewed; however, disregarding irrelevant information is nearly impossible. Just as a jury cannot un-hear evidence that has been stricken from the record (see, for example, Kassin & Sukel, 1997), it is difficult, if not impossible, to neglect potentially biasing irrelevant information obtained or reviewed as a component of a forensic evaluation. “Just as forensic examiners are well aware and take great steps to minimize physical contamination of the evidence, they also must be aware and take steps to minimize cognitive contamination. By focusing on the scientific data they need, and isolating themselves as much as possible from everything else, they minimize cognitive contamination and help achieve their role as scientists” (Dror, 2015, p. 4, emphases in original). As Neal and Saks (2016) highlighted, the decision task across different forensic evaluation types can be quite different; thus, the kind of information that is irrelevant, contextually biasing, etc. will be different for each forensic evaluation referral question and therefore the strategies used in an attempt to mitigate bias will likely vary by evaluation type or referral question.

Attempting to limit, as much as possible, the irrelevant information that is reviewed or considered as part of a forensic evaluation is one means of mitigating bias. Having a third-party take an initial pass through documents and records provided for an evaluation to compile relevant information for the evaluator’s consideration is one way of potentially mitigating against biasing irrelevant information (Neal & Saks, 2016). Another potentially mitigating strategy might be to engage in a systematic process of review where clear and specific documentation of what was reviewed, when it was reviewed, in the order in which it was reviewed, and with the evaluator detailing his or her thoughts, formulations, and inferences after each round of review, beginning with the most explicitly relevant case information (e.g., the police report for the index offense in a criminal responsibility evaluation) and moving toward the least explicitly relevant case information (e.g., elementary school records in a criminal responsibility evaluation). This iterative process of systematic documentation of data, inferences, logic, and conclusions would require evaluators to consider how their inferences and conclusions evolve as a result of various data sources considered and how these relate to the various hypotheses that are being tested. Wills (2008) described a related method of opinion formulation; however, as per Neal and Grisso (2014b), we caution that evaluators should be careful to engage in a systematic review that includes rival
hypothesis testing and documenting data and inferences that would both support and reject the various hypotheses being tested so as not to commit to an ultimate opinion prematurely. Hypothesis testing is an important component of the scientific method and the principle of falsifiability (see Popper, 1963) and several commentators have highlighted its utility and importance for forensic evaluation—confirming or disconfirming possible explanations for relevant behaviors and capacities (e.g., Grisso, 2003; Heilbrun, 2001; Melton et al., 2007; Milchman, 2015). Engaging in a process of systematic collection and consideration of relevant data to test alternate hypotheses, with careful documentation of what is being considered and how the various pieces of data either strengthen or weaken a particular hypothesis, can potentially mitigate the impact of bias in forensic evaluation.

**Reference materials**

Just as irrelevant case material can be biasing, so too can contextual information included in the reference materials for a forensic evaluation. Whereas in the forensic identification sciences reference materials pertain to the target or to a known suspect (e.g., the tire print from a specific make and model of car; the fingerprint of a known suspect), in forensic evaluation reference materials might include the legal test or standard for the referral question (including jurisdictional statutes and relevant case law), the abilities required of a defendant to meet a specific legal test, or any other metric that an evaluator is to use in considering whether or to what extent the evaluatee has certain characteristics or abilities, or meets the requirements for a legal test or issue in question. That is, reference materials would include whatever it is that the evaluator is supposed to be evaluating the evidence against and, of course, can include potentially biasing contextual information.

Admittedly, the decision tasks faced by forensic scientists differ greatly from those faced by forensic evaluators. Whereas forensic scientists are faced with epistemic challenges that convey a “right” answer (such as a known suspect) and their decision task includes an “identification” component, in forensic evaluation the decision task includes evaluating the extent to which a known individual’s characteristics, behaviors, or abilities meet a particular legal test, standard, or categorization. Just as fingerprint experts can “(re-)analyze the crime scene fingerprint mark after analyzing the reference print, with the risk that the expert starts to ‘see’ features that he [or she] would not have seen otherwise, and that might not be there” (Stoel et al., 2015, p. 76) so too can forensic evaluators become influenced by the reference materials that they consult for a forensic evaluation. Precedent, case law, and the interpretations of legal statutes and decisions can shape or impact our interpretation of the evidence in a forensic evaluation (see Philipsborn, 2004 for a discussion of how interpretations can be shaped by case law). Reference materials provide the basis for the legal issue being evaluated but can also influence interpretations or inferences made about the relevant data. Biasing contextual information within the reference materials may direct the interpretation of the actual data being evaluated. Such might be the case when the defense attorney provides information to the competency evaluator regarding the abilities required for the defendant to proceed to trial. Even though the information provided by a defense attorney regarding the abilities required for his or her client to proceed to trial is a necessary and relevant component of the reference materials for a competency evaluation, this information can influence an evaluator’s observations and inferences.

The reference materials also underpin the well-documented phenomenon of “rater drift,” wherein one’s ratings shift over time or drift from standard levels or anchors by unintentionally redefining criteria (see Kaplan & Saccuzzo, 2013). This means that evaluators should be careful to consult the relevant legal tests, statutes, or standards for each evaluation conducted and not assume that memory for or conceptualization of the standard or reference material is accurate. As one of the reviewers of this article noted, this point about rater drift appears to work well within the context of forensic mental health evaluation in light of Murrie and colleagues’ (2013) experimental findings about adversarial allegiance. Just as irrelevant or contextual information can influence our perceptions and interpretations, so too can our experiences and expectations influence our memories and conceptualizations, which might result in biased or inaccurate opinions or conclusions.

**Case evidence**

In addition to irrelevant case information and contextual information included as part of the reference materials for a case, the actual case evidence itself might also include some irrelevant, contextual, or biasing information. Here we conceptualize case evidence as information germane to the focus of the inquiry that must be considered by any forensic evaluator in arriving at an opinion about the particular legal issue. For example, police reports and discovery materials for the index offense in a criminal responsibility evaluation, or information regarding the allegations involved for a defendant’s criminal charges in a competency to stand trial evaluation. In essence, case evidence would include any information that must be considered in evaluating the particular legal
issue at hand (e.g., to proceed without the information would be considered inappropriate or sub-par practice by a majority of forensic evaluators). Whereas various background, educational, or treatment records may or may not be collected and considered in a criminal responsibility evaluation, to proceed without considering documented information regarding the index offense would be particularly problematic and generally viewed as sub-par or inappropriate practice for this type of evaluation. Influences at the case evidence level include biasing contextual information from the actual police reports or other data that must be considered for the referral question. Thus, contextual information that is inherent to the case evidence and that cannot be easily separated from it can influence and bias an evaluator’s inferences about the data.

Handwritten notes in the margin of an official police report or witness statement, the questions being asked of a suspect in a videotaped interview of a confession or other interrogation, or an officer’s description of a witness’ report can all include irrelevant or contextual information that can influence the perceptions and inferences of a forensic evaluator. Information that might be relevant to other experts in the case at hand or relevant to the trier of fact but that should be considered as impertinent information for the referral question being addressed by the forensic evaluator can be potentially biasing. Data, inferences, and conclusions about an appraisal provided by the retaining party, information provided by defense counsel in competency evaluations, victim impact statements in risk assessments, and official reports are all sources of relevant case information that can nonetheless be potentially biasing.

Irrelevant or contextual information can influence the way in which evaluators perceive and interpret data at any of these seven levels—ranging from the most basic aspects of human nature and the cognitive architecture of the brain, through one’s environment, culture, and experiences, and including specific aspects of the case at hand—but it is important to note that biased perceptions or inferences at any of these levels do not necessarily mean that the outcome, conclusion, or opinion will be biased. As West and Kenny (2011) make clear, bias direction and strength are independent. As explicated by Dror (2009), “If the bias is in the direction of the correct decision, then the evidentiary data may bring the decision considerations past the decision threshold, and the bias will only push the accumulated considerations further in the same direction” (p. 98). Even if the bias is in a contradictory direction from the correct decision, the evidentiary data might affect the considerations of the evaluator to some extent but not enough to impact the actual outcome of the evaluation or ultimate opinion of the evaluator. What appears important to the outcome is the degree to which the data are ambiguous; the more ambiguous the data, the more likely it will be that a bias will affect the actual decision or outcome (see, for example, Dror & Rosenthal, 2008). Ambiguous circumstances, therefore, are the most susceptible to biasing influences.

The path forward: Using science to improve forensic evaluation

In this article, we presented an overview of the various levels at which different factors can interfere with objective observations and inferences in forensic evaluation, potentially resulting in biased conclusions. Increased knowledge and understanding of the ways in which cognitive factors can impact a forensic evaluator’s perceptions and inferences is an important first step in attempting to mitigate the effect of bias in forensic evaluation; however, knowledge is not enough. The presence of a bias blind spot—the tendency of individuals to perceive greater cognitive and motivational bias in others than in themselves—has been well documented (see Pro-nin, Lin, & Ross, 2002). Neal and Brodsky (2014) demonstrated that forensic psychologists are occupationally socialized to believe that they can and do practice objectively (recall the discussion of training and motivational influences); however, emerging research on bias in forensic evaluation has demonstrated that this belief may not be accurate (see, for example, the discussion of adversarial allegiance). In addition, it appears that many forensic evaluators report using de-biasing strategies, such as introspection, which have been proven ineffective and some even deny the presence of any bias at all (Neal & Brodsky, 2016). What is clear, then, is that there is room for improvement in understanding and mitigating the effects of bias in forensic evaluation.

As Dror (2009) noted, “the need for improvement by itself is by no means a sign of a lack of scientific rigor; on the contrary, a sign of any good science is its constant reflection, criticism, and examination” (p. 94). For forensic evaluation to advance and improve, we must behave as scientists. “Scientists continually observe, test, and modify the body of knowledge. Rather than claiming absolute truth, science approaches truth either through breakthrough discoveries or incrementally, by testing theories repeatedly” (NRC, 2009, p. 112). Identifying weaknesses and taking actions and countermeasures to avoid or limit those weaknesses must underpin any science (Dror, 2009). Countermeasures such as Linear Sequential Unmasking (LSU; Dror et al., 2015)—where irrelevant information is masked, and relevant information is sequentially revealed when needed —can be adapted for use in forensic evaluation. Approaching
forensic evaluations like scientific inquiries and using rival hypothesis testing might place the necessary structure on the evaluation process to determine the differential impact of the various data considered (see Kassin, Dror & Kukucka, 2013 for further suggestions in the area of forensic science, and Dror, Kassin, & Kukucka, 2013 for debating the issues and objections).

In addition to using principles of the scientific method to improve the way forensic evaluations are conducted, these same principles can be applied to test assumptions and hypotheses about forensic evaluation. Identifying weaknesses in forensic evaluation and conducting research and hypothesis testing on proposed counter measures to reduce the impact of bias will serve to improve the methods and procedures in this area. Being scientific about forensic evaluation and using scientific principles to understand and improve it appears to be a reasonable path forward for reducing and mitigating bias. Three considerations will be briefly highlighted.

**Evaluator characteristics.** Recent research has demonstrated wide variability in the rates of evaluation outcomes as a function of evaluator, with some evaluators opining incompetence or insanity in a majority of evaluations whereas other evaluators almost never opine incompetence or insanity (see Murrie, Boccaccini, Zapf, Warren & Henderson, 2008; Murrie & Warren, 2005). The need for reliability among evaluators (as well as by the same evaluator at different times—inter- and intra-evaluator consistency) is a cornerstone for establishing forensic evaluation as a science (Dror, 2016). By understanding the characteristics of evaluators—including training, culture, and experience—that contribute to their opinions we can begin to propose and study different ways of limiting the impact of these characteristics on objective observation and inferences in forensic evaluation. Understanding and characterizing the abilities required for the specialized task of forensic evaluation will ultimately allow further development of profiles for selecting evaluators or for developing necessary evaluation skills.

**Evaluation methods.** Research has demonstrated that reliability improves when standardized inquiries are used for competence evaluation (see Grisso, 2003; Zapf & Roesch 2009). Preliminary research has indicated that the use of clinical interview data in conjunction with file review data in risk assessments leads to different rating outcomes regarding relevant risk factors than when only file review data is considered (see Kropp & Hart, 2000). Conducting systematic research on the methods and procedures used in forensic evaluation and the impact of these on evaluation outcomes and bias will ultimately allow for development of the most effective strategies for forensic evaluation. In addition, examining the interrelations between evaluator characteristics and evaluation methods will further increase our understanding of the interplay between personal and situational variables in forensic evaluation.

**Training.** Implementing professional training programs that address cognitive factors and bias in forensic evaluation and conducting systematic research on the impact of various training techniques for increasing understanding of these issues will likely improve the methods that forensic evaluators currently use to mitigate the impact of bias in their work (see Neal & Brodsky, 2016, for a discussion of strategies currently being used by evaluators). Indeed, Wettstein (2005) demonstrated that certification programs for forensic evaluators improve the reliability and quality of forensic evaluation reports and Gowensmith and colleagues (2015a) highlighted the importance of both initial as well as ongoing training. Although a trend toward state certification programs for forensic evaluators has been recently demonstrated, “most states still do not have a formal process for selecting or certifying” their forensic evaluators (Gowensmith, Pinals, & Karas, 2015b). In addition, a recent survey of forensic evaluators indicated that most evaluators are not provided any formal, direct training on the ways in which bias can impact their work (Zapf, Kukucka, Kassin, & Dror, 2017). Understanding the most effective ways of training evaluators to perform forensic evaluations in a consistent and reliable way while limiting the impact of bias will allow for the implementation of best practices, both with respect to the evaluations themselves as well as with respect to training procedures and outcomes. Applying scientific principles to test and compare various training strategies and outcomes will serve to improve knowledge and the practice of forensic evaluation.

**Summary and conclusions**

Consideration of the various influences that might bias an evaluator’s ability to objectively evaluate and interpret data is an important component of forensic evaluation. We have delineated seven levels at which various influences might interfere with objective observations and inferences in forensic evaluation. Evaluators are encouraged to be mindful of these various influences and to consider the use of various bias mitigation strategies, as noted throughout. Knowledge about the ways in which bias can impact forensic evaluation is an important first step; however, the path forward also includes the use of scientific principles to test alternative hypotheses, methods, and strategies for minimizing the impact of bias in forensic evaluation. Using scientific principles to continue to improve forensic evaluation will bring us closer to the
aspirational goal of objective, impartial, and unbiased evaluations.

References


