



Challenges to reasoning in forensic science decisions

Barbara A. Spellman^{a,*}, Heidi Eldridge^b, Paul Bieber^c

^a University of Virginia School of Law, 580 Massie Road, Charlottesville, VA, 22903, USA

^b RTI International, 3040 E. Cornwallis Rd., Research Triangle Park, NC, 27709, USA

^c The Arson Research Project, Monterey College of Law, 100 Col. Durham Street, Seaside, CA, 93955, USA

ARTICLE INFO

Keywords:

Forensic science
Cognitive bias
Reasoning biases
Similarity judgments
Causal attribution

ABSTRACT

The success of forensic science depends heavily on human reasoning abilities. Although we typically navigate our lives well using those abilities, decades of psychological science research shows that human reasoning is not always rational. In addition, forensic science often demands that its practitioners reason in non-natural ways. This article addresses how characteristics of human reasoning (either specific to an individual or in general) and characteristics of situations (either specific to a case or in general in a lab) can contribute to errors before, during, or after forensic analyses. In feature comparison judgments, such as fingerprints or firearms, a main challenge is to avoid biases from extraneous knowledge or arising from the comparison method itself. In causal and process judgments, for example fire scenes or pathology, a main challenge is to keep multiple potential hypotheses open as the investigation continues. Considering the contributions to forensic science judgments by persons, situations, and their interaction, reveals ways to develop procedures to decrease errors and improve accuracy.

Human reasoning has strengths and weaknesses. This article describes some characteristics of reasoning that could be of interest to forensic scientists. Among the goals is to identify ways to exploit and enhance the relevant strengths of human reasoning and to compensate for and ameliorate the relevant weaknesses. Ironically, among the strengths of human reasoning is that we automatically integrate information from multiple sources, which helps us find consistencies and generate causal explanations. However, good forensic analysis often asks that forensic scientists look at and evaluate some pieces of evidence independently of everything else that is known about a case; that is, it demands that forensic scientists reason in non-natural ways. Individuals and laboratories can develop procedures to facilitate accurate analysis even while imposing constraints on how it may be done.

Section 1 gives an overview of human reasoning, focusing on three characteristics that are especially important to forensic science. First, when reasoning, humans automatically take information from multiple sources and combine it to help us comprehend, use, and predict. The sources include not only the information in front of us, or that we learn from other people, but also our pre-existing knowledge and our motivations. Second, humans are “irrational” in that our reasoning often does not follow the formal rules of logic, probability, and statistics. However, the deviations we make are not random; rather they are systematic, showing that we often use reasoning shortcuts (heuristics) that will

usually, although not always, get us to the correct answer. Third, how well we reason depends on a combination of factors including characteristics of both the person and the environment, and also their interaction.

Section 2 describes the application of those reasoning processes described in Section 1 to forensics in general. It highlights how characteristics of human reasoning – both generally and specific to an analyst – and characteristics of the environment – both generally in a laboratory and specific to a case – can create problems for good forensic decision making. Identifying and acknowledging such problems can help lead to solutions.

Sections 3 and 4 consider challenges with respect to specific forensics. Section 3 looks at similarity judgments in feature comparison fields (e.g., fingerprints, firearms, handwriting, bitemarks). These disciplines can often remove external biasing influences from their decision making but other challenges may arise during the comparison process itself. Section 4 looks at causal and process judgments (e.g., fire debris, bloodstain spatter analysis, pathology). These forensic disciplines are typically searching for the story of how something happened rather than specifically who did it (although, of course, the analysis contributes to the search). Here, potentially biasing context information is often necessary for the analysis, however, other types of reasoning challenges arise, and a different set of debiasing techniques becomes important. Section 5 briefly concludes.

* Corresponding author.

E-mail address: spellman@law.virginia.edu (B.A. Spellman).

1. General features of human reasoning¹

Humans are great reasoners: We determine the causes of diseases and often eradicate them; we design and build huge structures in which we live and work; we develop and use sophisticated techniques to help us figure out who committed a crime. Humans are terrible reasoners: Most of us cannot multiply 5864×397 in our heads; we lose money playing the lottery, gambling, or in the stock market; and computers easily beat us in chess.

What are the characteristics of human reasoning that allow us to be both so good and so bad at it? Some of it is within us. Our minds automatically integrate information, seek patterns, and create causal stories. We are motivated to be correct but also be consistent. And we are profoundly, and usually unconsciously, influenced by the situation and environment we are in when we reason.

1.1. People automatically combine information

To make sense of incoming information (e.g., a sight, a word, an argument), people instinctively combine information from multiple sources. We create coherent stories from potentially unrelated events; we create categories for things we deem similar; and we construct our interpretations of information from a combination of what is in the world (“bottom-up” processing) and what is in our heads (“top down” processing).

1.1.1. Reasoning is both top-down and bottom-up

Much of the information we take in on a daily basis is ambiguous, incomplete, or overwhelming. Your officemate just said, “I went to the bank during lunch.” Did she get cash or enjoy a stroll along the river? You see a small cat. Is it really very small or is it just far away? The stimuli that we are exposed to every day could be interpreted in many ways but our various mental systems (e.g., sensation, perception, cognition) work together to interpret the stimuli in a sensible way. People are generally good at this kind of information integration (which may be why we enjoy being “tricked” by illusions).

Consider Fig. 1A. When asked which of the two horizontal lines is longer, most people answer the one on the bottom. But most people would be wrong. In fact, in the Müller-Lyer line illusion, the two horizontal lines are of equal length. This illusion relies on context – equal things may look bigger or smaller depending on what they are compared to (and we make those comparisons automatically). But the illusion also relies on pre-existing knowledge: people who live in more industrialized urban areas are more likely to fall prey to the illusion, and to experience the size difference as larger, than people who live in less right-angled, built-up environments [1].

In addition to perceiving unequal lengths, most people who see the illusion cannot “unsee” it, even after measuring the two lines. This phenomenon illustrates “cognitive impenetrability” – that sometimes even though we know something is true, we cannot make ourselves perceive it as true (e.g., [4]). These two features of reasoning: that our interpretations of information are affected by both top-down and bottom-up processes, and that sometimes our knowledge cannot de-bias our (false) perceptions, have important implications for forensic science.

1.1.2. People create abstract knowledge structures: categories, scripts, and schemas

A second important way in which we combine information is by taking individual instances of things – objects, people, events – and

¹ Most introductory and cognitive psychology textbooks will cover these relevant characteristics of reasoning – but the coverage will reside in many different chapters including: Perception, Memory, Categorization, Reasoning, Social Cognition, and Judgment and Decision-Making. In this overview section, references are provided only to seminal papers and highlighted experiments.

incorporating them into abstract general knowledge structures. Categories are a ubiquitous type of general knowledge structure; categories contain objects that we know are different but we treat as the same for useful purposes [5]. We have categories like trees and dogs (“natural kind categories”) and like chairs and xylophones (“artifact categories”). And we have created categories like blood type, fingerprint class characteristics, and illegal drugs.

Categories are useful for memory (e.g., I can say “I saw 8 birds outside today” rather than trying to remember each one). They are also useful for inductive reasoning (because if something is a member of a category, we expect it to have characteristics common to the category). The tree will be tall and green; the dog will bark; the chair can be sat upon. But as with most generalizations, category features don’t always apply. For example, most people nod agreement to the statement “birds fly” even though we know that not *all* birds fly (penguins, ostriches) and not *only* birds fly (some insects, bats).

Learning categories takes experience and/or instruction. There needs to be a reason (a rule or knowledge about the category members – implicit or explicit) for which characteristics of objects (i.e., “features”) matter to putting them into the same category. In forensic science, novice analysts need to learn what features of potential evidence matter in order to understand the relevant categories in their science (see [6]; this issue, in general; e.g., [2]; with respect to fingerprints).

Although categories are immensely useful for understanding and communication, things can go wrong in several ways: when people are trained on a set of items that are not representative of the category (e.g., when people learn the wrong base rates [7]); when people do not know the extent of the variability within a category; when people weight features incorrectly when determining which category items belong to; when the learning/training does not match later testing environment; and when people have pre-existing beliefs about the rules for categorization. Categories also can be resistant to modification (See Section 3 discussion of similarity and [6]; this issue.).

Other types of general knowledge, which typically aren’t talked about as often as categories, are “scripts” and “schemas”. Scripts are general outlines of how to behave in situations that we view as sufficiently similar to previous situations, for example, how to behave at a new fast food restaurant or doctor’s office. “Schema” is the broader term for general knowledge structures; it and includes those above plus others. For example, many people have schemas for different types of robbery – likely gleaned from watching too much bad television. For a bank robbery, people may imagine several robbers wearing masks, shouted instructions for everyone to lie down, guns, a note handed to a teller, a fast escape, and a waiting getaway car (among other things). When research participants listened to an audio recording of a bank robbery trial and were asked to report what witnesses said, they sometimes injected events from their own schemas that were not actually reported, for example, they might misremember that people were threatened even though they had not been [8].

Thus, scripts and schemas help us understand single events by grouping them with other similar events. Like categories, we use them to make inductions: If it is a bank robbery there must be a threat. Generalizations like this can be good because they can help prompt our memories to retrieve what might have occurred; however, they can be bad because they may fill in our malleable memories with incorrect information [9].

1.1.3. People create coherence: Story model and explanatory coherence

Scripts and schemas describe how we use generalizations from similar repeated event sequences. This section describes how we might interpret a novel set of events. When individuals hear information that could be about a connected set of people and occurrences, we automatically (and typically unconsciously) try to fit all the information together into a causal story that makes sense and can account for all of the information. Mostly we succeed. But sometimes we get it wrong.

Two lines of research illustrate this general process. The “Story

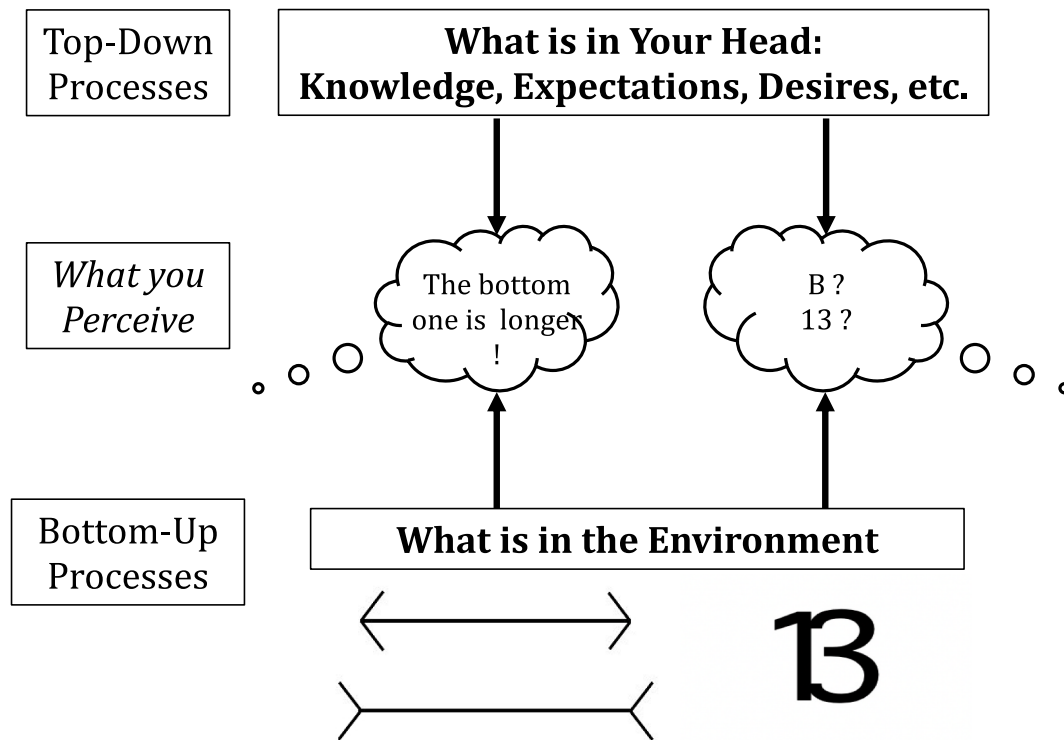


Fig. 1. 1A (left) and 1B (right). Interpretation depends on both top-down and bottom-up processes. 1A is the Muller-Lyon illusion (see [2]; for more discussion); 1B is from [3]; and discussed in Section 2.

Model” [10] examines how individuals, acting as jurors, come up with verdicts after hearing information from a trial simulation (transcript or video). The general important finding is that people try to take all the facts and fit them together into one story that explains everything, including what happened and why people behaved that way. (Because this very specific task is most relevant to causal and process judgments, it is described further in Section 4.)

The broader line of research, called “Explanatory Coherence” is consistent with the Story Model but creates a wider framework, and a connectionist computational model (ECHO [11], for understanding reasoning from evidence and, in particular, how people choose between competing explanations [11]; for a legal example see [12]). It describes how the decision about what makes the best interpretation of events depends not only on the amount of supporting evidence, but also on the simplicity of the explanation, the number of alternative explanations, and its consistency with (or analogy to) other beliefs [13] – values often mentioned in philosophy of science (e.g., simplicity, parsimony). Suppose a prosecutor has an eyewitness who testifies that he saw the suspect at the crime scene; the prosecutor will now increase her belief that the suspect committed the crime. Later she learns that a latent print from the suspect was found at the crime scene. Now, not only does the prosecutor increase her belief that the suspect committed the crime, but she also increases her belief that the eyewitness was correct. Because each consistent fact supports, and is supported by, other consistent facts, people treat them as non-independent; thus, explanatory coherence illustrates why it is difficult to change any one belief (e.g., that the eyewitness identification was faulty) once there is other consistent information. These characteristics of reasoning have important implications for both comparison judgments (esp. the dangers of exposure to potentially biasing information) and causal processes judgments (esp. the dangers of not creating or mis-evaluating competing hypotheses).

1.2. People are often said to be irrational

Despite the many good decisions people make, much research describes human reasoning as being irrational. In the 1960’s through 1980’s, a great deal of cognitive psychology research illustrated human irrationality in the sense that judgments and decisions do not always adhere to the formal rules of logic, probability, and statistics.

One example of irrationality is that people fall prey to the “Gambler’s Fallacy”, believing, for instance, that after a roulette wheel lands on red several times in a row, the wheel has a greater chance to land on black than it would have otherwise. Many such illogical beliefs were demonstrated by Amos Tversky and Daniel Kahneman [14]; see [15] for a collection of the classic papers), who made the important point that although people’s reasoning deviates from formal rules, the deviations are systematic, not random, and thus could give researchers insight into how reasoning is done. So how are people reasoning systematically but not rationally? Kahneman, Tversky, and others argue that reasoners use heuristics, which are short cuts in reasoning that usually get us to the correct answer but may lead us to a wrong one. For example, the availability heuristic tells us that things that come to mind more easily are more common than things that come to mind less easily. This heuristic usually works. Are there more people in the US named Lancelot or John? A quick mental check of the people you know gives you the correct answer. But try to estimate how many English words have the form: _ _ _ N _ . Now try to estimate how many have the form: _ _ _ I N G . When different groups of people are given those patterns, the estimate for the latter tends to be much greater than that for the former, even though the former must (logically) fit more words. But it is so much easier to imagine words that fit the _ _ _ I N G pattern.

In addition to pointing out the potential irrationality of heuristics, these researchers also began identifying biases in human reasoning. Like

heuristics, biases can be useful in some environments and harmful in others. Biases that are particularly important in forensic analysis, and methods for debiasing, are discussed in Section 2 below.

1.3. Explaining the “irrationality”

If people don't come up with the logical answer, yet mistakes aren't random, what are they doing and how are they doing it? One explanation of how heuristics work is that when people are asked a question that is too difficult to answer (because, for example, there is not enough time, the information is not available, or we do not have enough mental capacity), we substitute an easier question for the difficult one. For example, instead of creating a full count of Lancelots and Johns, we think of the people we know, assuming (rightly or wrongly) that it will be a representative sample for the names. If we see two piles of coins on the floor and one has more coins than the other, but we do not have time to add up the values, US adults assume that the one with more coins is worth more than the one with fewer (the “numerosity heuristic” [16]); even though we know that we could be mistaken.

1.3.1. Why? System 1 versus system 2 reasoning

The Cognitive Reflections Task (see Fig. 2) provides some good examples of heuristics at work. Each question is fill-in-the blank and has one correct answer. In theory, people could give a wide variety of answers. However, in practice, for each question most people give one of two answers – either the correct one, or, of the infinite number of possible wrong answers, one specific wrong answer [17]. Why?

The phrasing of the problems suggests certain answers. Consider problem 1. We hear the pattern of 5/5/5 and then the beginning of the pattern 100/100/? Humans are great at finding patterns, which are often, but not always, useful in reasoning. Why shouldn't this pattern continue with 100? And, in fact, people give the answer 100 quite frequently, even though the correct answer is 5. The phrasing of problem 2 also suggests certain answers. The question mentions doubling so the answer should involve cutting something in half. Many people then quickly divide 48 by 2 and get the (incorrect) answer of 24 days. They would get the correct answer if they did what your 5th grade teacher told you: “carefully check your work.”

The Cognitive Bias Codex [18] is a useful (and huge) graphical taxonomy of heuristics and other types of influences on reasoning. It puts almost 200 biases into categories; the organizing categories have to do with limits of memory, overwhelming amounts of information, lack of meaning in some tasks, and the need for fast decision making. These are exactly the types of situations in which our minds need help to do the right thing.

A useful way of thinking about reasoning, and when and why it might go wrong, is to consider that there may be two different cognitive “systems” at work. Dichotomies in human reasoning have been proposed for many years along many dimensions [19]. An overarching view, which provides a convenient way to think about reasoning (even though the theory is still up for debate), is to think of one system as being “intuitive” and the other as “reflective”. System 1, the intuitive system, has the characteristics down the left side of Fig. 3. Every day, we take many actions and make many decisions automatically and unconsciously. We react to hearing our name across a crowded room in a conversation we were not attending to; we drive our regular route home while obeying the traffic laws, listening to the news, but without remembering that we meant to stop for milk; we change our behavior in response to environmental cues that we cannot report. System 2, the reflective system, is slower and takes intentional mental effort. We cannot do our tax returns at the same time as playing a game of chess; each demands our full cognitive resources to be at work.

Most people could find the correct answers in the Cognitive Reflections Task, but we do not because we go with the first sensible answer that comes to mind. That is, we rely on the answer provided by the heuristics offered by the quick and intuitive System 1, rather than

slowing down and intentionally engaging System 2. Of course, this is not to say that System 1 is always wrong (far from it) and System 2 always correct (far from that as well), but for some types of problems it is worth slowing down to check (more on this in the next section).

Importantly, both our intuitive- and reflective-based decisions may be influenced by factors that we are not aware of and without recognizing that we are being influenced. People are quite willing to point out that other people are biased in how they come to a conclusion; however, we are quite unable to recognize such biases in ourselves. This “bias blind spot” results (in part) because we cannot “feel” biases operating within ourselves; we feel as if we are only operating with the information we are conscious of and as if we are capable of only using the relevant and ignoring the irrelevant information [22]. Acknowledging that people (including ourselves) can be influenced without knowing it, and that we might not be able to ignore information even if we are aware of it (e.g., the examples in Fig. 1), are first steps in recognizing the importance of having techniques for debiasing.

1.3.2. Who and when? It depends on the person, the situation, and their interaction

It seems that you can fool some of the people some of the time but how often depends on both the people and the time. There are some personality traits that correlate with making fewer errors on reasoning problems. One is Need for Cognition [23] – which measures how much a person enjoys and chooses to engage in effortful thinking. When the original study using the Cognitive Reflections Task (CRT) compared how often students at different universities made errors, it found that fewer errors were made at MIT (Massachusetts Institute of Technology) than at Princeton, Harvard, Michigan, and other top universities [17]; p. 29). Follow-up studies argue that people who do better on the CRT are less susceptible to reporting quick answers without further reflective analysis – it is not because they lack intuitions [24].

Another important factor about people that matters to performance on such tasks is their expertise on the types of task presented. Daniel Kahneman and Gary Klein [25] seemed to have a long-running debate about whether intuition is stupid (according to Kahneman) or smart (according to Klein). What they realized was that – it depends. Kahneman was studying novice decision makers; Klein was studying experts. They agreed that someone who has worked in a predictable environment in which they had made similar decisions frequently and gotten good reliable feedback (i.e., had developed expertise), would develop good and reliable (and fast) intuitions for problems within that environment. Thus, experts' intuition can be very good (as exhibited by very good performance under speeded conditions by fingerprint examiners, although slowing down improves performance; [26,27]). Yet, as described later, even experts may make mistakes because of biases.

But characteristics of individuals are not the only things that matter to decision making; the environment in which judgments are made is also important. System 2 reasoning takes time and cognitive resources. When people are rushed to make a decision, they are more likely to rely on System 1 reasoning than System 2 reasoning. When people are engaged in multiple tasks, so that they cannot pay full attention to either, they are more likely to rely on System 1 reasoning. When people are operating at the edge of their knowledge, they may rely on System 1 if they cannot rely on System 2. For example, there is much worry that when jurors do not understand complex scientific testimony they may rely, instead, on superficial (“peripheral”) cues to decide which of two experts to believe, such as the experts' credentials. This effect has been shown for mock jurors, especially those low in Need for Cognition [28] and for real jurors who heard testimony in homicide trials [29].

However, to reiterate from above: It is not the case that System 1 is always wrong and System 2 is always right. Very often, they come to the same conclusion. Experts' System 1 can be very accurate. And it has been suggested that for some types of complicated decisions, System 1 may be better than System 2. But in forensic science, at some point, it is probably best for people to slow down and let their System 2 check their

- | |
|---|
| <p>1. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes</p> <p>2. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days</p> |
|---|

Fig. 2. Subset of the cognitive reflections task [17].

System 1 (Intuitive)	System 2 (Reflective)
Fast	Slow
Automatic	Controlled
Nonconscious	Conscious
Heuristic	Logical
High capacity	Capacity limited
Parallel	Serial
Contextualized	Abstract
Independent of cognitive ability	Correlated with cognitive ability

Fig. 3. One variation of the System 1/System 2 distinction [3,20]).

Note. The well-known distinction between peripheral and central processing, can be viewed as a dichotomy along these lines with the low-cognitive engagement of peripheral processing seen as System 1 and the more intense engagement of central processing seen as System 2 [21].

work.

2. Reasoning in forensic science

Section 1 described some of the natural human tendencies in reasoning.² But forensic science often requires analysts to reason in non-natural ways. Why? The output of a forensic analysis is supposed to be free from biases and independent from the output of other investigative processes. That is, analysts are supposed to reason only from their general scientific knowledge, their expertise, and the evidence in front of them, not from any task-irrelevant information about the situation or the case.³ The integration with other evidence or case information is the responsibility of different personnel for different tasks (e.g., police for investigation; lawyers for evaluating how strong the case is and how their side should proceed; judges and jurors for ultimate verdict decisions). Thus, for analysts, the analysis process is intentionally decontextualized from some of the clues that System 1 might use to trigger heuristic reasoning. Instead the process should rely heavily on System 2 reflective reasoning. However, like System 1, System 2 also responds to potentially biasing information.

2.1. Sources of reasoning errors

Despite its connotation, “bias” is not necessarily a dirty word: it need not be intentional and it need not get you to the wrong answer. A useful definition of cognitive bias is:

The class of effects by which an individual’s preexisting beliefs, expectations, motives, and situational context may influence their

collection, perception, or interpretation of information, or their resulting judgments, decisions, or confidence.⁴

Note how the definition includes biases that are built into our reasoning apparatus which, most of the time, like heuristics, help us get quickly to a correct answer. It also includes outside sources of information that might influence our judgments, or our confidence in our judgments, for the wrong reasons – even though the influence might be in the correct direction.

Fig. 4 divides influences on forensic (or any) judgments vertically into two major types: Person and Situation (sometimes called “Environment”). This divide, as an explanation, comes from a book by the influential psychologist Kurt Lewin [35]; who tried to reconcile the debate between those who argued that all human behavior could be explained by either one or the other; instead, he claimed that behavior was a function of both. The figure is also divided (horizontally) into levels of generality: what can be attributed to people or situations generally versus what is particular to specific people or situations. The quick current bottom line is that characteristics of a person, and of a situation, each matter to behavior, but so does the interaction between them, including the effects that they may have on each other and the “fit” between them [36].⁵

2.1.1. General human reasoning: Motivated reasoning, confirmation bias, tunnel vision, and bias blind spot

The key to understanding bias is this: What people know, believe, learn, and want can influence how they search for, interpret, evaluate, and integrate information (see Fig. 1). And these influences can come

² Other articles and chapters describe the relevance of some similar and some different qualities of the human mind to forensic science, e.g., [2, 30–33].

³ What is considered task-relevant “other information” depends on the discipline and tends to differ between feature comparison and causal process disciplines.

⁴ This definition is mostly from [34]; p. 45, but modified by the first author after conversations with Itiel Dror.

⁵ This figure owes much to the “triangle” figure of [32]; but it pulls apart the factors to illuminate the independent contributions of person and situation, which should be useful when considering how different types of interventions could be used to remedy problems.


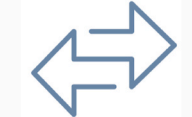
	GENERAL	SPECIFIC
	General Human Reasoning <i>Applies to (almost) all people</i>	Analyst-Specific Reasoning <i>Varies across people</i>
Person 	Create Coherence and Seek Patterns Use Heuristics Motivated Reasoning Confirmation Bias; Tunnel Vision Bias Blind Spot	Individual Cognitive Characteristics (Including Talent and Ability) Individual Personality Characteristics Specific Prior Experience (e.g., Statistical Learning, Stereotypes) Beliefs and Preferences (Prejudices) Education and Training
	General Situation Factors <i>Varies across laboratories/disciplines</i>	Case-Specific Situation <i>May vary by case</i>
Situation 	Laboratory/Discipline Expectations Laboratory Procedures Treatment of Errors Allegiance Effects	Pressures (e.g., Time, Outside Expectations) Comparison Processes and Reference Materials Exposure to Case-Irrelevant Information ("Contextual Bias")

Fig. 4. Characteristics of Analysts and Situations that may affect decision making.

Note: This representation includes much of the same information as the “triangle figure” in [32]. Of that paper’s eight sources of bias: Person/General includes items from his #8; Person/Specific includes #4, 6, and 7; Situation/General includes #5; Situation/Specific includes #1, 2, and 3.

from many sources – both internal and external. But, as mentioned above, “bias” is not necessarily a dirty word. Many biases are built into our reasoning apparatus and, most of the time, like heuristics, help us get quickly to a correct answer. However, because of the highly influential power of forensic results – from police investigations, through plea bargains and trials, to the appeal process – the criminal justice system would prefer that forensic science get to the correct outcome all of the time [37].

Cognitive biases can be distinguished in a number of ways. One is along the dimension of whether the bias pushes reasoners towards particular desired outcomes or not. The general processes of seeking patterns, creating coherence, and using heuristics, as described in Section 1, by themselves, are content neutral – that is, it does not matter what is being reasoned about. Research has shown, however, that when reasoning, individuals may be motivated by, and affected by, several things including: getting to the correct answer, appearing consistent, and getting to a particular desired answer; these desires comprise “motivated reasoning” [38]. Because such desires inhabit our top-down reasoning processes, they can affect judgments differently depending on content.

Fig. 1B illustrates a stimulus from an experimental task [3] in which participants see a series of pictures; some participants are told to press a button if they see a letter whereas others are told to press a button if they see a number. For each identification they get correct they will get points toward a reward. A series of letter and numbers appears and the participants get nearly all of them right. However, when they see Fig. 1B, people in both groups often press the button. Being correct is important. But when things are “close” or ambiguous, people’s motivations may push their perceptions in the particular desired direction.

This type of “motivated reasoning” appears in several guises in forensic analysis and across other aspects of forensic work. For example, forensic scientists who work in a police laboratory, rather than an independent laboratory, might be (consciously or unconsciously) motivated to help the police catch perpetrators by more often interpreting ambiguous information consistent with what they believe the police want. Although forensic scientists are often heard to claim that they are completely impartial, practices required by many laboratories such as granting solo pre-trial conferences to the prosecution but requiring prosecutorial notification of pre-trial conference requests by the defense

reflect a systemically biased culture of which forensic scientists may not be consciously aware, but which may nonetheless influence their reasoning. Or, in the verification stage of an ACE-V analysis (Analysis, Comparison, Evaluation, Verification), someone might be more likely to verify the determination of a friend or superior than an analysis by someone unknown or disliked or might more closely scrutinize the work of a disliked colleague that they hope to catch in an error. In a stunning experimental demonstration of motivated reasoning (in this case, an “allegiance” or “role” effect), 108 forensic psychologists and psychiatrists were paid to review the same offender case files and score the offenders on two risk-assessment instruments. On average, the participants who were told that they were consulting for the prosecution rated the offenders as riskier than the participants who were told that they were consulting for the defense [39].

A too-common related bias is confirmation bias. Confirmation bias is in action when one already has a pre-existing belief and validates it by seeking out new evidence to support it, interpreting ambiguous evidence in favor of it, and discrediting or overlooking information that disfavors it [34,40]. Confirmation bias occurs in law in many guises; for example, when the police have a suspect and then seek information that will nail him but at the same time they drop the investigation of all other potential suspects. Failing to pursue alternative suspects or theories, and only valuing information that is consistent with what one already believes is also known as “tunnel vision” (or “mind-set” in friction ridge comparison; [41]).

And a huge problem, for novices and experts alike, is the bias blind spot – that people see bias in others but not themselves [22]. A survey of over 400 forensic science examiners revealed that most believed that cognitive bias was indeed a “cause for concern” in the forensic sciences as a whole (71%), fewer that it was a concern in their specific domain (52%) and fewer still that their own judgments were influenced by cognitive bias (26%). This latter number is perhaps most concerning because it demonstrates that relatively few examiners acknowledge the possibility for bias in their own casework and even if it exists they believe they could set aside biasing influences. For example, on average, they believed that they sometimes know what conclusions they are expected to find, but not that such knowledge would affect the conclusions they reach [42].

2.1.2. Specific human reasoning: personality, knowledge, training

Peoples' minds differ from each other in a myriad of ways, as a result both of inherent individual characteristics and particular lifetime experiences. Reasoning success may depend on personality characteristics such as "grit" ("conscientiousness") and Need for Cognition (which predicted success on the CRT in Section 1).

Reasoning success will also depend on the knowledge and experience analysts have with the topic at hand. Expert forensic analysts have been through training and practice before they are allowed to enter the profession; they then have individual experiences as analysts. For some disciplines, there are also differences in underlying talent and ability that may carry through training and eventual practice [43]. (For more about individual traits and types of training see [44]; and [6]; this issue, respectively.)

Among the important factors that can bias individual decisions is the statistical learning that an analyst has had – that is, how many different examples has the analyst seen in her experience and how often has she seen them (and does she have other information about their frequency)? Acquiring such information is important in developing expertise – and the intuitive competence of experts (as in [27]). Such statistical information teaches people to determine which features of a stimulus are more or less important and indicative of some conclusion [31]. However, as with the use of categories and schemas, assuming regularities can affect judgment, for example, familiar case information (like similar case reports) can lead novices to make the same, but sometimes incorrect, decision on a subsequent analysis [7] and irrelevant information that triggers a stereotype can affect pathologists' decisions [45,46].

2.1.3. General Situation: laboratory procedures and expectations

The "General Situation" quadrant of Fig. 4 applies to laboratories or disciplines (or other group settings) in which a number of people are treated similarly – and perhaps differently from other groups. Laboratory procedures and expectations may affect reasoning about all cases in a variety of ways. For example, if analysts work in a police laboratory, they could show "allegiance effects" – being biased to give an answer they perceive that police want to hear.

Laboratories might also work under more or less pressure and impose more or less stress than other laboratories, thereby also affecting judgments. A stressful laboratory, demanding relentless fast work, could encourage more fast heuristic judgments resulting in errors. Analysts, with different expertise and different personality profiles, may react in a variety of ways – including burnout, or in the extreme case, fabrication of results ([47,48]; this issue).

Laboratories and disciplines find some types of errors to be worse than others – and may have different consequences for analysts if they make an error. For example, laboratories may expect that analysts should never make an error or they may be more critical of some types of errors (e.g., an incorrect identification) than others (e.g., an incorrect exclusion). Consequences may include having to complete root cause analyses, having previous and current casework reviewed, being temporarily removed from casework or subjected to remedial training, or in extreme cases, termination. Such expectations could affect analysts' decisions, especially in close calls. Fingerprint analysts have been shown to be "conservative" in making identification judgments – that is, their errors are more in the direction of inconclusive or exclusion judgments rather than toward identification judgments [49].

Laboratories can also make it easier, or more difficult, for analysts to learn potentially biasing task-irrelevant information about cases they are working on. For example, in some laboratories analysts will always learn the type of case it is, the names of the people being investigated, and other kinds of "contextual" information. But because forensic laboratories engage many people working on many cases doing many tasks, laboratories can sometimes be a major catalyst for improvement. For example, some larger laboratories are able to separate the analyst performing the work from a case analyst who evaluates the case

information to determine what work is to be done.

2.1.4. Case-specific situation: task-irrelevant information

There are many important factors relevant to case-specific work including time and other pressures on reasoning (mentioned in "general situation" above) and factors within the structure of the task performance itself (see Sections 3 and 4 below). However, much of the work on forensic reasoning in the last 20 years has been concerned with the effects of "mental contamination" [50] -- being exposed to "task-irrelevant" (or "contextual") case information, which is not relevant to a decision and should not be needed to make it accurately, but, if learned, may affect it.

One type of irrelevant yet biasing information is what other analysts have determined. Early and dramatic illustrations came from research by Dror and colleagues [51,52]. In one study, 6 fingerprint examiners were each shown eight prints that they themselves had previously judged as individualizations or exclusions. The prints were presented with contextual information that suggested that the prints should be evaluated as the opposite as they had been in the past (e.g., a previous exclusion was paired with information that the suspect had confessed – suggesting individualization). With such biasing information, the analysts changed their decisions on 17% of the trials. Subsequent studies across disciplines, and studies of verification processes, have also shown such influences (see Sections 3 and 4 below).

More common than learning about what another analyst would decide with the same information is the problem of learning task-irrelevant case evidence. Suppose an analyst is doing a comparison of fingerprints or bullet striations or tool marks. She then learns that the suspect who owned the tool or firearm or finger had confessed to the crime that created the crime scene from which the evidence was taken or she learns the outcome of another relevant forensic analysis. With that knowledge, she turns back to her comparison. This type of mental contamination can create several problems. First, it could cause her to make the initial evaluation that there was enough information in the latent print to be useful for analysis rather than rejecting the print on those grounds. Second, as described above, it might push an analyst towards a particular decision she would not otherwise have made. Third, if it is consistent with the analyst's pre-existing leanings, it will increase her confidence in her decision. This so-called "corroboration inflation" [53] or, more generally, "forensic confirmation bias" [34], is a natural outcome of our reasoning system (recall "Explanatory Coherence" from Section 1.1.3). Indeed, if an analyst's finding is consistent with other (accurate) findings, then it should be more accurate. However, the size of the uptick in confidence tends to be unwarranted, thus making it excessively more difficult to appropriately devalue the total weight of the evidence if one piece is shown to be erroneous later on. The fourth problem is that when an analyst aggregates information, it ultimately takes away from the jury or judge the availability of *independent* evidence on which to make a decision. Jurors are supposed to assume that the various evidence they hear was arrived at independently; when they use information that already takes into account other evidence, the factfinders will be, in a sense, double-weighting the contaminating evidence [37].

Similarly, when a causal and process analyst (e.g., fire, bloodstain pattern, pathology) learns details of the case not directly related to the analysis, based on the Story Model he will likely – intentionally or not – integrate it into the causal story he is creating [10]. As with comparison evidence, the mental contamination might push an analyst into an explanation he otherwise would not have made; if consistent with the analyst's leanings, will increase his confidence in his explanation; and will ultimately take away from the jury or judge the availability of independent evidence on which to make a decision.

Note that consistent with Explanatory Coherence, how much the contextual information affects a judgment should depend on various qualities of the information and the informant(s). Information that is repeated becomes more believable and weighted more heavily [54]. And

information that comes from more than one source is valued more than information that comes from only one source. This makes sense: information that is corroborated by a variety of people should be, in fact, more likely to be true than information that is not. However, such information would be devalued if the analyst has the cognitive resources to recognize the sources are related (e.g., they likely had gotten the information from the same place; [55]). Information that comes from particular informants should be, and often is, appropriately devalued. People use information about a source's history of lying and their professed certainty when divulging past information to assess how much to believe the source at present [56]. And, people give more weight to information that they believe others are trying to hide from them [57, 58].

2.1.5. Interactions

The four quadrants in Fig. 4 consider people and situations separately. But as Lewin noted, behavior is a function of both person and situation, and their interaction. Not all people will react to all situations in the same way. For example, forensic examiners vary widely in how much stress they report and what it is due to (e.g., caseload/backlog or perceived management expectations). The amount of stress reported also varies by forensic discipline and experience [47]. Acknowledging that various factors are at play can suggest whether interventions to improve individual or lab performance should be specific or general in scope.

2.2. Preventing, undoing, or debiasing potential reasoning errors

We just described several general challenges to reasoning in the forensic context. They illustrate ways in which there may be some deviation from the pure, analytic, and implausibly perfect reasoning processes desired by forensic science. Can the thought processes that lead to an error be repaired? Can reasoning be de-biased? To some extent yes. Different techniques apply to ameliorating different causes of error. But in general, to improve performance, the best path is to change the decision environment [59].

The processes described in the General Human Reasoning quadrant, which are built-in to our reasoning apparatus, are basically immune to change. As laid out in Section 1, people automatically combine information, “see” coherence, and support our own side in arguments. Fifty years of studies on experts in different domains (including medicine, accounting, sports, chess, music) show that developing expertise does not necessarily provide protection against such processes, even when known to be unwarranted [60]. The problems that arise particularly from Analyst-Specific Reasoning can be ameliorated in part by analyst selection (see [44]; this issue) and in part by training and creating common knowledge among analysts (see [6]; this issue).

The contribution to errors from the situation should be thought in terms of being a result of the interaction between how people reason and the circumstances in which reasoning takes place. Laboratories can create procedures for making the General Situation more amenable to good reasoning (see [48]; this issue). But Case-Specific Situations, and the interaction between analysts and case-specific situations will continue to create more nuanced challenges.

We next describe general steps that can be taken to prevent bias. The most important thing to learn is that debiasing is tough to do and, if possible, it is better to find ways not to introduce biases in the first place than to try to overcome them later.

2.2.1. Before analysis

The best way to stop biased or bad reasoning is not to be subjected to conditions that promote it. Of course, for things that are either inherently parts of our reasoning systems or immutable parts of the environment we work in, that cannot be accomplished.

On the other hand, some biases that are created through mental contamination might be preventable. For example, some forensic

laboratories have implemented case management systems in which only a case manager is aware of all aspects of what comes to the lab (e.g., location of the crime scene, which types of forensic tests are being done, etc.), and analysts only know information essential to their own testing (e.g., the material that a latent print was taken from; storage conditions of a substance; material a bullet was removed from). Such a system relies on analysts not also being investigators on the same case.

Keeping extraneous information away from analysts is called “blinding.” Blinding – or intentionally keeping information away from a decision maker – is a common technique. In a double-blind drug study, neither the doctor nor the patient knows whether the patient is receiving the novel treatment or a placebo. In blind peer-review, reviewers do not know who authored the paper they are evaluating. In forensic science, blinding means limiting analysts to information that is relevant to the task the analyst is expected to perform [61]. Of course, sometimes biasing information may be inextricable from the evidence that forensic analysts must examine. For example, a handwriting analyst examining a threatening written note cannot help but read the content of the note, which could be biasing. Bloodstain pattern analysis often reveals information about the nature of the crime. In such cases, blinding would not be an option and debiasing techniques might be engaged later.

A survey of 183 forensic analysts across different disciplines revealed that most of them believed that some kinds of case information were not relevant to their tasks (e.g., information regarding a suspect's or victim's name is not relevant to their evaluation of firearms). This finding suggests that having case managers who blind analysts to all case knowledge could be implemented without much objection. However, there was great variability across individuals and disciplines [62,63]; see more discussion of discipline differences in Sections 3 and 4 below.

2.2.2. During analysis

Several different types of strategies can be used to promote better reasoning during the analysis stage. One is to make sure that the analyst has enough time and available cognitive resources to do an analysis that relies on System 2 rather than only System 1. Another is to provide external analysis tools or procedural aids (described in the sections below). But what if an analyst has been exposed to some “mental contamination”? Can the effect of the information be eliminated by instructions to do so during the analysis? This problem has been extensively studied in the analogous context of jurors hearing information that they are later told to disregard.

Popular television courtroom dramas provide great examples of mental contamination: a lawyer asks a question, a witness blurts out an answer, the opposing lawyer yells, “objection”, and the judge tells the jury to “disregard” what was said. Millions of viewers rightly wonder how the jurors could possibly disregard whatever highly important information they have just heard. Note that the jurors' job is not to forget the information (although it would be useful if they could); rather, it is to make their decisions as if they had never heard the information in the first place. The hundreds of studies of disregarding show that various factors may affect whether people will do so – and raise the question whether people are unable to do so or unwilling to do so [58,64]. Some factors relate to the evidence: type of evidence matters (e.g., people will disregard unreliable more than reliable evidence) and how integral the evidence is to the case matters. This last point is related to Explanatory Coherence: Information that is linked to other information, especially that which supports beliefs based on other information, or that ties information together into a story, probably cannot be “extracted” without having changed how the other information is understood and used [54].

An additional context in which getting rid of biasing information has become important is in the dissemination of “misinformation” and fake news. Many techniques have been tried to debias people; only a few ways to reduce the impact of misinformation have worked. One is a pre-exposure warning: telling people that what they are about to hear may be misleading with an explanation of how it might mislead them. Second, is if people are told time after time to “retract” the information. And

third, and best, is to retract the information but replace it with an alternative narrative, including an explanation of why the original information was mistaken in the first place [54].

A technique that has promise for reducing the effect of biasing information is the creation of “forensic science line-ups” – that is, having analysts look for “match” to a standard from a set of items rather than comparing one-to-one [65,66]. Reasons for and against such procedures are described in Section 3, after an explanation of how biases may arise because of comparison processes.

2.2.3. After analysis

After an analyst makes a (potentially biased) judgment, and after methods to de-bias a particular judgment are employed, there are ways to reduce the probability that the system overall will end up with a biased judgment. One is to get other people within the forensic lab involved. For example, as already in use in the ACE-V approach to some of the feature comparison disciplines (i.e., fingerprints, shoeprints, firearms, toolmarks) analysis, a second analyst might be called on to verify the work of the first – although the process varies widely across labs [67]. The usefulness of this technique will depend on what the verifier knows about the details of the case, whether the verifier sees notes from the initial analysis, and whether the verifier knows the identity and decision of the initial analyst. When firearm examiners used blind verification (i.e., did not see the initial examiners notes or proposed conclusions), they were about 5 times more likely to disagree with the proposed conclusion than when they did view the initial examiner’s notes and proposed conclusion [68], thus suggesting strong confirmation bias in verification.

2.3. Errors, biases, and consequences

Analyses of wrongful convictions show that forensic science makes mistakes. The Innocence Project [69] reports that of their first 375 successful exonerations, 43% involved misapplication of forensic science. However, despite the pervasive concern with biases, biases do not necessarily lead to errors. In fact, sometimes biases will lead one to make a correct judgment. But, as explained below, regardless of accuracy, biases can lead to problems further down the line.

2.3.1. A biased judgment need not be an incorrect judgment

Although mental contamination is clearly a problem, many of the ways in which its biasing effects have been studied have revealed that it is only a problem in “close” cases – that is, in cases in which the existing evidence could reasonably support differing conclusions. For example, when real fingerprint examiners see the same pair of prints they have evaluated before, but this time with contaminating information that suggests the opposite conclusion [51], such potentially biasing information affects the ultimate decision only when the relevant evidence does not heavily support one conclusion over the other. The “motivated reasoning” literature (Section 2.1) describes how people may be motivated to come to a particular conclusion, but they are also motivated to be correct. Thus, even when people are biased to want a particular outcome, when the evidence is clear, that bias will succumb to the desire to be accurate.

Learning task-irrelevant information might also get one to the correct answer but for the wrong reasons. For example, suppose you are a fingerprint examiner and you learn the results of a DNA test before doing the comparison. Your evaluation of the fingerprints is likely to be biased by the DNA conclusion. Curley et al. [70] describe such a study (but with novices [71]) and argue that although sometimes the DNA biased the participants in the wrong direction (when the DNA analysis was wrong), sometimes it made the participants more accurate (when the DNA analysis was correct). They claim that “cognitive biases may actually improve accuracy.” However, as [72] points out, you got to the correct answer but for the wrong reason and you are overconfident in your conclusion. Biases can work in either direction, but your fingerprint

interpretation is not independent from the DNA interpretation, and police, judges, and jurors cannot properly evaluate it [37]. Even worse, suppose that a court determines that the DNA was inadmissible, or further analysis reveals the trace DNA was a mixture, or the analysis was incorrect. The tainted DNA results are infiltrating the trial (via your testimony), now with no possibility of being scrutinized under cross-examination. (See also [73]; regarding context information improving pathologist’s interpretation of patterned skin injury, and the critique by [74].)

2.3.2. Noticing and responding to errors

Will potential errors be discovered and then repaired? One problem with noticing potential errors is that, as described previously, people, including forensic analysts, do not recognize when their own reasoning may be biased (the “bias blind spot”; [22]), so they believe that no “repair” is needed.

A second problem with responding to errors is that errors often go unnoticed because there is either no feedback or the feedback is circular. By the former we mean that no one (except a verifier, if one is used), ever comments on the accuracy of the output. Of course, because in real forensic cases ground truth is not known, accurate feedback does not exist (for the importance of accurate, timely feedback, see [6]; this issue). But much worse than no feedback is circular feedback. A detective one of us knows once asserted, “I have never elicited a false confession. After confessing, the suspect always either took a plea deal or was convicted in court.” Obviously, what the detective failed to take into account is that the confessions were likely causes of the verdict (or plea deals) and that he had no independent measure of the suspects’ actual guilt. The same is true of forensic science. It is such powerful evidence that knowing the results may be the major cause of a suspect taking a plea bargain or a jury deciding to convict [37]. Thus, such outcomes may be effects of the forensic determinations, not independent evidence verifying whether they were accurate or mistaken.

A third problem is that it is easy to explain away errors in hindsight, to come up with mistaken explanations for errors, and to believe that “the same thing” could not happen again. All of these impede learning from, and appropriately responding to, errors. For example, suppose a lab makes a mistake. Perhaps two samples (e.g., drugs, DNA, latent prints) got mixed up when being checked into the lab and they are attributed to the wrong case file. When discovered, an analyst might say, “Oh, I knew something looked fishy when I saw the log. Next time someone will notice and say something, so it won’t happen again.” That is an example of hindsight bias: people claim that they had known it all along – and would notice it next time [75].

Alternatively, a lab member might say, “The mix-up is Nick’s fault. He was fired, so it won’t happen again.” That is an example of attributing an error to a person who is a “bad apple”—blaming qualities of a person rather than qualities of a situation (known as “the fundamental attribution error”). Maybe Nick was especially careless but maybe there was some case specific factor (e.g., six items came into the lab at once) or the general way that samples are handled that would make it better to consider the mistake as related to the lab procedure rather than to an individual. After getting rid of Nick, the lab would feel more positive about its processes because Nick was gone but should not because the underlying problem was not addressed.

Blaming analysts as a first pass is common (e.g., in the Mayfield and McKie cases), but the above mistake hypothetical is similar to what happened in 2001 in the Las Vegas Metro Police Lab to Dwayne Jackson, whose DNA was (wrongly) said to match DNA picked up at a robbery scene. As a result, he took a plea deal and spent nearly 4 years in prison. It is believed that two DNA samples were switched during technical processing and the switch was the result of human error. The forensic scientist on the case was put on leave and, sensibly, the lab also began a review of their quality assurance standards [76]. Investigations and repairs should consider people, processes, and environment, not only the specifics of the one particular situation [77].

2.4. Forensic discipline differences

Section 2 has described the sources of some possible errors and biases in forensic analysis and the difficulty of de-biasing. Some laboratories have adopted procedures that should eliminate some of the sources of bias (e.g., case managers) or that can ameliorate the consequences (e.g., verification by a second analyst). More specific types of potential biases and fixes are discussed in the following two sections on Specific Applications: Section 3 for Feature Comparison Disciplines and Section 4 for Causal and Process Judgments. However, regardless of the specific forensic discipline, it is important to remember that debiasing is tough to do and, if possible, it is better to find ways not to introduce biases in the first place than to try to overcome them later.

3. Specific Applications: feature-comparison – similarity and individuation judgments

The feature-comparison disciplines vary along many dimensions. The PCAST report [78] evaluates these methods: DNA (both simple and mixed), bitemarks, fingerprints, firearms, footwear, and hair. Other feature comparison methods that this section should be relevant to include: handwriting, documents, tire tracks, tool marks, and controlled substances.

Humans compare things all the time. We are good at finding identical things; we even create games that encourage children to “spot the differences” between two pictures. But forensics that involve feature-comparison are especially difficult because analysts are not looking for identity; rather, they are looking at whether two artifacts are *similar enough* to make some judgment about them – and that judgment is usually about whether they were generated by the same source.

In forensics, what is typically being compared are samples from a process – and like all generative processes, and sampling processes, there will be variability in the items. Plus, in these disciplines the generation mechanism itself is typically not examined and sometimes it does not even physically exist. Consider firearms. The analyst is often comparing a bullet that was picked up at a crime scene with one that was fired from the firearm being investigated at the lab. Although the question for the analyst is whether the bullets came from the same gun, the inside of the gun, need not be examined. What is compared are the artifacts created by the firing process. Fingerprint analysis is similar. Analysts do not look at a suspect’s fingers; they look at what is produced from two different sampling situations – one created in the wild and one usually created in a controlled situation (e.g., by police). Of course, sometimes the two samples are both from the wild, for example, two different crime scenes. Regardless, the firearms and the fingers do exist, limiting the potential variability.

What variability does exist can be exploited during analysis when analysts compare crime scene evidence to multiple standards. For example, firearm analysts typically test-fire more than one cartridge before making a comparison. If the firearm does not mark well, or marks inconsistently, they may test fire more cartridges. Even after the markings are imaged and microscopic comparisons are done, they may choose another cartridge (or two) to use for comparison based on whether it is clearly showing the marking they want to use. Similarly, fingerprint examiners may look through many exemplars of a suspect’s on-file prints (i.e., if they have many) to find the one that has the clearest reproduction of the area needed for comparison to the latent print. One problem with this procedure is that judgments of “clarity” depend on previous exposure and familiarity with similar (or identical) objects (see [79] for a review). Analysts therefore could be, unknowingly, selecting the closest matching test fire or standard for comparison. Another problem with looking at many exemplars is that examiners may shift the criteria for what “counts” as sufficient similarity by reinterpreting ambiguous features as more (or less) similar to the sample, as illustrated in Fig. 6 [80].

Superficially, analysis of handwriting seems the same as that for

firearms and fingerprints: the analyst compares two samples of handwriting, trying to evaluate whether they were generated from the same source. But where is the standard – the “origin” or prototype – from which both samples were generated? It does not physically exist and what is generated can be faked – intentionally trying to make it look similar to or different from other handwriting. Hair and DNA are different from those processes, but similar to each other, in that the samples come from an already-existing, though temporally changing, population of cells and hair. The samples (gathered, rather than newly produced), will also vary although, in theory, very little for DNA.

3.1. Seeing and evaluating similarity

Of course, humans make similarity judgments in our everyday lives and we recognize things as being generated from the same source, even when what we perceive is not identical. For some things we are quite good, like recognizing our friends in the street, even though today they do not look identical to any memory we have of them from any other time (new clothing, sunburn, bad hair day, etc.). Our good friends we recognize anywhere. However, when we see a mere acquaintance at the market, we might not be quite sure who she is; yet when we park next to her where we do our volunteer work together, the context makes the identification easy. Thus, like every other judgment that people make, comparison judgments are influenced by both bottom-up and top-down processes.

3.1.1. Biases before comparison

Section 2, above, describes several ways in which biases – mostly of the form of motivated reasoning (i.e., wanting a decision to come out a certain way) or of learning external information that suggests a particular answer – may push people toward conclusions not fully supported by the evidence. A conceptual example was from [39]; in which forensic psychiatrists made different assessments of offenders depending on whether the psychiatrists were told that they were working for the prosecution or the defense. However, Dror, Scherr, et al. [46] failed to find allegiance effects in a study with Forensic Document/Handwriting Analysts⁶ although such analysts have been shown to be susceptible to other contextual biases (e.g., [81]).

A classroom demonstration based on a case from Evidence Law [82] illustrates how seeking certain types of information can bias decision making. Law students viewed a schematic drawing of two explosive devices (see Fig. 5.). Half the students were asked to list similarities between the two devices (as the prosecution had done); the others were asked to list differences (as the defense had done). Those who had listed similarities were more certain that the same person had built the two devices than those who had listed differences. Thus, even though they did not start with different beliefs, the approaches they used to examine the same evidence affected their conclusions. In the real world, the approach that they took – looking for similarities or looking for differences – would likely have been motivated by their desires to support a specific outcome. Of course, the outcome of the analysis would then, in turn, strengthen their initial beliefs.

3.1.2. Comparison processes: which features matter?

The most important issue in assessing similarity is keeping this question in mind: Similar with respect to what? For example, which pair is more similar: (1) United States and China, or (2) United States and Italy? The answer differs depending on whether the question is about nuclear capabilities versus about the prevalence of olive oil.

In the *Trenkler* experiment, when asked to list similarities between

⁶ Reasons for not finding an effect are provided by the authors; these include the variability in answers, the artificiality of the stimuli, and the minimal instructions to the participants about whether the client was the prosecution or defense.

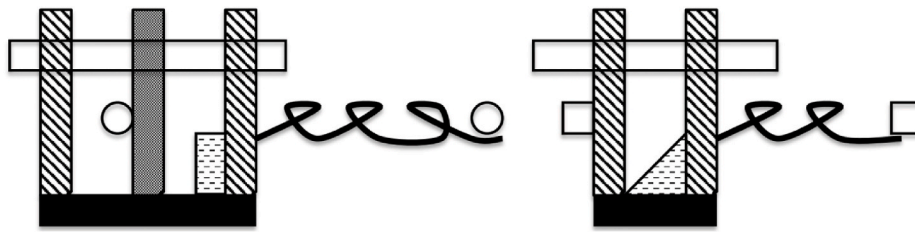


Fig. 5. Created by Spellman (nd).

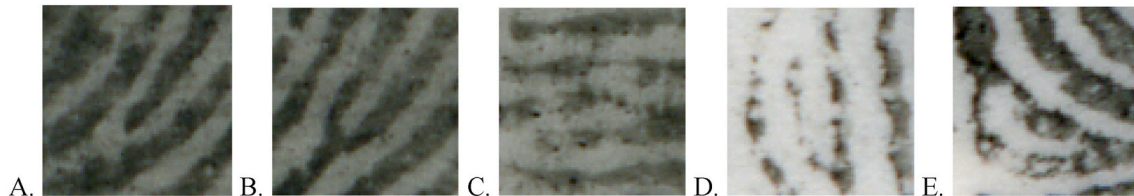


Fig. 6. Some simple and compound features in latent print examination. A – Ridge Ending, B – Bifurcation, C and D – Hooks, E – Enclosure, or Lake.

the two devices, people say things like: It has a black base, a white crosspiece, and vertical bars. There is a dotted piece supporting the bar that a curly fuse is attached to. The end of the fuse has the same shape object as one placed to the left on the device. People who are asked to list differences will describe how there are two versus three vertical bars (and the middle one differs in shading); the support pieces are different shapes; the fuses differ in number of curls; and the little pieces are round on one device but square on the other.

These descriptions illustrate how “what counts as a feature” is a subjective decision. Is having vertical bars a feature? Or is each bar itself a separate feature? And are attribute features (e.g., something is white) different from relational features (e.g., the small piece and fuse piece are the same shape)? And at what level of detail do we consider features? So, assume that having vertical bars is important. But is their height? Width? Material that they are made of? Do we count those all separately or are those one “thing”?

And what about correlated features? One device has two round pieces and the other two square pieces. Is the existence of the pieces, and their shapes, each a separate feature? What if the pieces are typically sold or used in pairs because they need to be of the same size and shape for the device to work properly? If knowing about one object guarantees (or even highly implicates) the form of another, how much weight should each independently contribute to a similarity evaluation?

As astute readers (and some participants in the experiment) have intuited, another important thing in assessing overall similarity, and each feature’s contribution to similarity, is how typical the features are in the population. Suppose there are many explosive devices of this type that are built by thousands of different people. And suppose all of them have bases and crosspieces and vertical bars. Noting that these two pictured devices are similar in that they each have those particular parts should do nothing at all to increase one’s belief that they were built by the same person. On the other hand, if most devices do not have any curls in the fuse, then the fact that both do here, regardless of the difference in number of curls, should contribute greatly to the similarity evaluation. However, balanced against this, feature comparison examiners must also keep in mind that discrepancies in appearance are often sufficient to support an exclusion.

Examples of the importance of these considerations come from across the range of feature-comparison disciplines. With fingerprints, examiners first consider “Level 1” features like the pattern and ridge flow (see Fig. 6). But, for example, for a fingerprint classified as a whorl, having two deltas is not a separate feature; it is part of the definition. Examiners may next examine Level 2 features, which are features in the path of a ridge, such as a place where a ridge ends or splits in two (i.e.,

bifurcates). Each of these events is typically referred to as a single “feature,” yet when two appear in close proximity, such as a bifurcation right next to a ridge ending (commonly known as a “hook”) or two bifurcations facing each other (commonly known as an “enclosure” or a “lake”), these form what are often referred to as “compound features” – does each of these compound features count as one feature, or as two?

For these reasons, as well as due to a lack of clear definitions of what constitutes a feature, there is a high variability among fingerprint examiners on how many “features” are present in a given image, even when the image itself is very clear [83,84]. Some forensic organizations used to require a specific number of matching features before a positive identification could be made. Although some organizations may still want a minimum number, it is clear that pure counts of similar/dissimilar features, with each weighted the same, makes no sense for most similarity judgments; more subtlety is necessary [85].

The bottom line is that when judging similarity, what features we see, which we think are important, and how we count and weight them, should depend not only what is in front of us but also on the question we are trying to answer. That is, like most human reasoning, it involves both bottom-up and top-down information. Identifying relevant features, knowing the prevalence of features, recognizing typical covariates and appropriately devaluing them, understanding how and why they were generated, are all part of what distinguishes experts from novices. So being able to ignore the influence of unimportant features [7,31]. People can make superficial similarity judgments quickly but even for experts to truly answer how similar are things with respect to a particular question requires time, mental resources, knowledge, and a good dose of System 2 (reflective, slow, logical) reasoning to check System 1 (intuitive, fast, heuristic) reasoning. Processes for evaluating comparisons should incorporate these factors.

3.1.3. Comparison processes: what changes during comparison?

As described above, which features will be noticed, and how much they count in evaluating similarity, depends not only on what physically exists, but also on the evaluator’s motivation and pre-existing knowledge. Another thing that matters, even less obviously, is how the evaluator makes a comparison.

New or different features may emerge when comparisons are made. For example, when experimental participants were asked to list similarities between Items A and B in Fig. 7, they wrote that each had three prongs. Other participants, who were asked to list similarities between Items C and B, wrote that they each had four prongs. Which is correct? The interpretation of that knob on the right side of Item B depended on what B was being compared to. Fingerprint examiners sometimes report

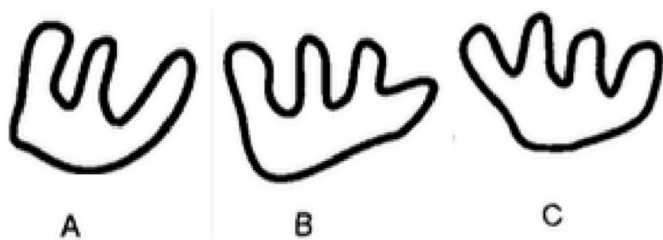


Fig. 7. From [86].

seeing more features (i.e., minutiae) when a latent print is presented on its own than when presented with a comparison print [83].

Not only does which items are compared matter but the direction of comparison may matter as well. There are consistent findings of similarity asymmetries for visual stimuli. Deformed shapes are rated as more similar to their prototypes (e.g., think ellipse to a circle) than the prototypes are to the deformed versions; simple figures are rated as more similar to larger complex figures of which they are a subset than vice versa [87]; there are also conceptual examples.

Finally, data from a slightly different kind of study suggest that the order in which features are considered and compared could affect judgments of similarity. In several marketing studies, participants were asked to hypothetically choose between two backpacks (or restaurants or coats). The two brands were rated as overall equally attractive, and rated equally good on some features, but they differed in desirability of other features (e.g., construction, material, zippers, price). When a brand was described with its best feature first and its worse one fourth (of six total features), it was selected as the more desirable brand 2:1 over the other brand which had been described with its worst feature first and its best feature fourth. That is, even though people started out with no bias between the two brands, and learned analogous information about them, the order of learning about the different features mattered [88]. Studies like this one are consistent with studies showing confirmation bias. It suggests that when making similarity judgments, early discoveries of similarity would be weighted more than later discoveries of differences (and vice versa).

3.2. Minimizing unwanted effects

The above problems – like the biasing effects of task-irrelevant information and that comparisons, and direction of comparison, may change perceptions of features and of evaluation of similarity – have been of concern to PCAST and other observers. Laboratory procedures such as ACE-V (Analysis, Comparison, Evaluation, Verification for fingerprints; [89]) and LSU (linear sequential unmasking; compare [90], to [91]) techniques are designed to ameliorate or eliminate some of those problems in several ways.

For example, for most feature comparison processes, much of the time, potentially biasing extraneous information can be kept from an analyst by a case manager, at least at a first pass. However, the content of a handwritten message or, for instance, the necessary disclosure of what substance a footprint or fingerprint was lifted from, or what substance a bullet traveled through, might provide biasing information. ([32]; calls this source of bias “data”.) Feature comparison analysts believe much of the evidence commonly found on evidence submission forms is task-irrelevant. Forty practicing forensics analysts, who attended one of five trainings and self-identified with a primary discipline that could be considered “Pattern Evidence,” participated in the survey of analysts mentioned above [62]. They answered questions about their perceptions of the task-relevance of 16 types of case information that are commonly found on submission forms [63]. The pattern evidence analysts were very consistent (at least 80% agreement) that nearly all information about the suspect and victim (including things like names, ages, race/ethnicity, and criminal history), plus the names of the

investigating officer, eyewitness accounts, and offense type, were task irrelevant (11 items). The information most wanted by the analysts was a description of the evidence (76%). Analysts were more split on the relevance of information about the method of evidence collection, case synopsis, and suspect’s statement and alibi.

If analysts believe that information is task irrelevant, and if the information is of the sort that could be potentially biasing, it seems that there is no good reason to provide it to them. This research does, however, come with certain caveats. First, there was never 100% agreement among the analysts about the task-irrelevance of any of information. This lack could be explained by differences across individual examiners or across the disciplines combined into the Pattern Evidence category. Second, the overall sample is rather small, preventing finer grained analyses, and, because they were at a training, might not be representative of all Pattern Evidence analysts.

In addition to blinding analysts to task-irrelevant information, ACE-V and LSU incorporate other lab procedures to eliminate other types of potential bias – especially with regard to the order of procedures. Analysts first examine and document the trace evidence alone before comparing it to the standard. This process should (a) create a record of the analyst’s initial observations; (b) move the analyst from System 1 to System 2 reasoning, and (c) make the analyst aware of emerging or changing features when the analyst makes the comparison to the reference sample [41].

A possible change to comparison procedures in the case of exposure to biasing materials could be the addition of “fillers”. Analysts could be asked to do the equivalent of police line-up – selecting a suspect (i.e., a “matching” item) from several different items, with all but the suspect’s evidence from known-innocent sources. A fingerprint study with novices shows that this procedure improves discriminability compared to the standard comparison procedure when there is incriminating contextual information [66]. However, expert examiners in a study that did not include contextual information showed only very small (mostly non-significant) differences between the two procedures [65]. Both articles note that this procedure would be time consuming but it might have other benefits such as revealing fraudulent analysts.

Verification of the decision by a second analyst should ameliorate the effects of any idiosyncratic biases that the initial analyst might have. There is some debate, however, about what a verifying analyst should be told. Being told of the initial analyst’s decision and seeing the initial analyst’s notes could create a bias -- and has been shown to do so in firearm examiners [68]. A simulation by [26]; based on data from fingerprint analysts, shows that having multiple independent analysts, and using majority rule for the decision, would eliminate most errors. However, having verifying (or additional) analysts start from scratch could be quite time-consuming (with unknown ultimate value according to [91]).

4. Specific Applications: causal and process judgments (e.g., fire scene examination, bloodstain pattern analysis, pathology)

Unlike in the feature-comparison disciplines described above, in the disciplines that involve causal and process judgments, analysts do not have evidence in front of them to compare to a known standard. Rather, they have evidence and want to develop hypotheses, using knowledge and experience, about what likely caused that evidence to exist in its current form. For example, given the details in the destruction found in a building fire, analysts want to determine the area of origin of a fire, what materials were first ignited, and what heat or ignition source caused the fire to start. Bloodstain pattern analysts want to know what things, movements, and sequence of events created a particular visual pattern. Pathologists want to know cause of death.

At a deep level, in terms of the human reasoning processes involved, these tasks have many similarities to feature comparison tasks. However, because of the variety of information that causal and process (“C&P”) analysts are exposed to, and the types of judgments they are

supposed to make, some potential problems in reasoning loom larger. Below we consider potential difficulties based on how evidence is obtained, on how initial hypotheses are created, and on how hypotheses are updated and revised. In C&P analyses, it is very clear that both bottom-up and top-down reasoning processes are implicated in judgments.

4.1. Obtaining information

The information that C&P analysts need to create an explanation does not magically appear to them completely, all at once, and neatly packaged. Analysts may be part of the discovery process (e.g., conducting an autopsy, going to a scene) and obtain different pieces of information at different times. The information they get may be incomplete, may contain irrelevant information that cannot, on the surface, be distinguished from relevant information, and may even contain intentionally misleading information as when a criminal adds, distorts, or eliminates useful evidence. All of these make the analyst's task to determine the correct explanation more difficult.

4.1.1. Potentially biasing information from the investigation

Like feature-comparison analysts, causal and process (C&P) analysts may learn about analysis-irrelevant evidence in a case. Because of their investigative roles, the exposure seems quite ordinary, and it is often unclear which evidence is or is not relevant to their judgments.

Unlike most feature comparison analysts, C&P analysts often go to the scene where evidence is being collected. With evidence that is large – for example, an entire building burned or blown up, scattered debris, bloodstain patterns across walls and floors -- it is simpler, and maybe even better in some ways, to bring the analyst to the evidence rather than the evidence (or pictures of the evidence) to the analyst. Of course, during such an excursion, it is easy to be exposed to other case-relevant but analysis-irrelevant information from the site itself or from other people involved in the investigation. In fact, fire investigators routinely interview witnesses before going to the fire scene or at the fire scene to get clues as to where to begin a search and to begin forming hypotheses about where and how the fire began. Such information might include, for example, memory or speculation about what the scene looked like prior to the event, routines of people who might have been involved, what electrical appliances were or were not in use, etc. [92,93]. This information, especially if received early in the investigation, could initiate confirmation bias in the analysis for how the event occurred. Plus, it violates the objective that the pieces of evidence presented at trial be independent from each other.

Even forensic pathologists, investigating cause of death, may be biased by seemingly innocuous medically-irrelevant information. In an experiment [45,46], 133 board-certified pathologists read a vignette about a case in which a caretaker brought a 3 ½ year old child to the hospital with a skull fracture and brain hemorrhage; the child died soon after. The pathologists all read the same descriptions of the scene and the medical examination, but half read that the child was African-American and was brought in by the mother's boyfriend while the other half read that the child was White and was brought in their grandmother. Most pathologists ruled the manner of death as "undetermined", but of the pathologists who determined a cause, those who read the boyfriend version were more likely to rule the death as a homicide rather than an accident whereas those who read the grandmother version ruled it was more likely to be an accident than a homicide. Note that whether such bias is a function of individual base rate or stereotypes, removing its effect is probably easiest by changing lab procedures.

In short, C&P analysts are, often by necessity, exposed to a lot of potentially biasing context information – sometimes intentionally, sometimes not; sometimes task-relevant, sometimes not (e.g., bloodstain patterns: [94,95]; fire: [96]).

4.1.2. Potentially biasing effects of their role

C&P analysts' role in investigations may also have a biasing influence on their generation and evaluation of hypothesis. Not only might they know that they are working, directly or indirectly, as part of the police department or prosecutor's office, but, unlike most of the feature-comparison determinations, C&P analysts' final determinations are often conclusions about whether a crime has even occurred. For example, fire investigators are often asked to take the next investigative step and to conclude whether the fire was ignited by accident or on purpose; a determination that a fire was started with an accelerant dispersed throughout a building screams "crime". A forensic pathologist may be asked to determine whether a wound could have been self-inflicted – to distinguish potential suicides from homicides. A similar question about bloodstains could be posed for bloodstain pattern analysts. Thus, not only might C&P analysts be feeling pressure (consciously or not) to find what the police want, but C&P analysts might also be even more motivated toward a specific finding because their analyses speak to whether there is a crime to be investigated at all. Indeed, fire investigations are often called "arson investigations" and C&P analysts often seem more like crime scene investigators than discipline-specific analysts.

4.2. Creating initial hypotheses/explanations

As described back in Section 1, the human mind is quick to combine information; we find patterns and create coherent stories – sometimes even when such patterns or coherence does not exist.

4.2.1. Existing evidence

We mentioned "The Story Model", which is based a line of research in which "mock jurors" read or hear the facts and testimony from simulated criminal trials, and are asked to choose one verdict from a set of possible verdicts. Different studies varied features within the trials. One important finding was that people try to take all the facts and fit them together into one story that explains everything, including what happened and why the characters behaved that way. For example, in a trial, Johnson, the defendant, is charged with first degree murder. He and the victim had a fight earlier in the day and, in a later fight that evening, Johnson pulled out a knife and killed the victim. However, whether the killing was intentional, accidental, or in self-defense was unclear. Participants who believed that Johnson was angry and went home during the day to get the knife were more likely to think it was an intentional murder; people who believed that someone like Johnson might normally carry a knife, and that the victim threatened him, were more likely to think it was self-defense. Note how the differing verdicts are products of not only the evidence presented, but also of participant's pre-existing knowledge and expectations about typical human behavior [10].

These studies show that people prefer to create stories that include as many of the presented facts as possible. They also prefer stories that are "coherent" – including being consistent and plausible [10]. Related studies show that for a set of facts, people prefer simpler explanations to more complicated ones and single explanations to multiple ones (e.g., [13]). Of course, these preferences might compete for a given set of facts. Fitting facts into one story results in the devaluation of facts that do not fit into the story. After selecting a verdict in the Story Model, participants later remembered the trial information that was consistent with the verdict they selected better than the inconsistent information – suggesting that they had devalued that inconsistent information.

Devaluing some trial testimony or some evidence picked up at a crime scene is not necessarily an error – testimony may be mistaken or false; evidence may have been planted or misinterpreted. However, the challenge comes later when new evidence comes in and a C&P analyst has to be open to alternative hypotheses about the case and those hypotheses relay on those facts that were devalued because of confirmation bias.

4.2.2. Missing information

C&P analysts may also face the problem of missing information. In fire or explosion analyses, original material and patterns may be obscured or completely gone. In bloodstain pattern analysis, it is unknown how many people were involved in its production.

As described previously, humans are pattern-seekers and coherence-seekers, we create stories and use schemas – for better or worse – to fill in the blanks when information is missing. Missing evidence, of, for example, an expected bullet hole in a wall, could be a product of the hypothesized bullet not existing, or of the bullet being deflected in a way that did not leave the expected discoverable trace. The analyst might think “no hole, no bullet” or might recite the maxim “absence of evidence is not evidence of absence” and assume there had been such a bullet regardless. The work on story construction suggests that once C&P analysts have already been exposed to a lot of case information, and have a leading theory of the case, they are likely to fit the hypothesized missing evidence into that already-existing theory. They might be correct; however, the lack of such evidence could be a critical piece of an alternative correct hypothesis. It is important to consider the probability of such a negative finding under each hypothesis and use that information when revising hypotheses as an investigation continues [97].

And as mentioned previously, research on jurors hearing inadmissible evidence [64] and on the staying power of misinformation [54] suggests that the more important the hypothesized information, and the more integrated it becomes into the case explanation, the more unlikely it is that analysts will be able to discount or disregard it when subsequently trying to constructing alternative explanations.

4.3. Generating and evaluating multiple hypotheses

As new information comes in, C&P analysts should be keeping their minds open to generating new hypothesis and re-evaluating older ones. Generating alternative hypotheses can be difficult; people tend to be drawn to a narrow set of hypotheses based on limited information and conventional thinking [98], and then often stick with earlier alternatives even as more information is discovered (consistent with confirmation bias or tunnel vision). However, people can be prompted to consider alternatives [59] although some analysts will be better at this than others. There are individual differences in how likely people are to converge on a single story to explain a case versus keeping multiple theories open as more evidence is gathered [99](see [100]; for a description of foxes versus hedgehogs). There are also cognitive tests (e.g., Need for Cognitive Closure [101]), that can help reveal who is, and is not, inclined to jump to conclusions; see [44]; this issue).

After generating multiple hypotheses, investigators must decide which ones are plausible and which one provides the best explanation (and how good those explanations are). In large part, fire investigation views getting to the right analysis as a process of generation followed by elimination. The analyst creates alternative hypotheses and then begins eliminating them by gathering more evidence. In an odd and controversial twist, when one hypothesis remains, it will often be accepted, even in the absence of any evidence that supports that hypothesis directly. For example, once the location of the origin of a fire has been established, an analyst might then consider the various potential causes of the fire (e.g., cigarette, electrical wiring, appliance). After ruling them out, the analyst concludes that the ignition source was a match or a lighter that was removed from the scene after igniting the fire. Note that this determination is not based on physical or empirical evidence; to the contrary, it is based on an absence of evidence [93].

That description reveals several problems of the elimination technique. First, we know that deductively, when every possible account of the evidence is generated as a hypothesis, then if all but one is ruled out, the remaining one, however implausible, must be true (at least according to Sherlock Holmes [102]). However, in real life generating all possible explanations is typically impossible, so that deduction would not be valid. People can easily generate plausible hypotheses to fit

existing information but which ones they will generate and which they will test may be influenced not only by appropriate information but also by other top-down information, including prior beliefs, knowledge, and motivations. Second, the evidence under evaluation might not be 100% reliable. Witnesses' memories may be incorrect about who they saw where and when. Burn patterns – just as blood spatter and anatomical irregularities – might be consistent with more than one hypothesis, but with different likelihoods. The most useful evidence about a fire may have been erased by the fire itself. Plus there is always the possibility (as in intelligence analysis) that someone planted or changed evidence with an intent to mislead the investigation. Finally, this process of elimination provides a ready-made shortcut for an analyst to draw a pre-determined conclusion regardless of the evidence or lack thereof.

Overall, given the potential unreliability and incompleteness of information, and the necessity to both generate and evaluate hypotheses, C&P analyses are complex tasks that provide many challenges, both bottom-up and top-down, to good decision making.

4.4. Minimizing unwanted effects

The major things to look out for in C&P analyses are biases in the initial generation of potential explanations (through role biases, exposure to potentially biasing information, and the use of schemas) and biases that affect the ability to appropriately generate and evaluate multiple additional hypotheses.

Laboratories could reduce exposure to some potentially biasing information by assigning case managers (as for feature comparisons) but removing all contextual information might be impossible or counter-productive for C&P analysis. C&P analysts not only are typically exposed to context information, but also want to learn and use it. The survey on analysts' beliefs about what is task-relevant information [63] did not contain a category specifically for C&P judgments. (The categories were: Pattern Evidence, Forensic Biology, Chemistry, and Crime Scene Investigation). The 12 analysts who said their primary discipline fell into the crime scene investigation category were the most divided about what counted as task-irrelevant information and did not have over 70% agreement in calling any of the 16 types of information “task irrelevant”. (They also were, by far, the smallest sample, although 38 analysts noted it as a secondary discipline.) Experimental research with 39 bloodstain pattern analysts attending a training conference found that all of them were interested in receiving some kinds of contextual information (e.g., medical findings, police briefing, DNA results). Receiving the information, especially the medical findings, biased the analysts' judgments [94, 95].

One suggested way for reducing biasing effects of contextual information in bloodstain pattern analysis is to have an independent context-blind “checker” for the analysis (as in [68]). A New Zealand forensic agency designed a protocol in which such context-blind checkers received diagrams of the items containing bloodstain patterns and, if they wanted, limited information about, for example, the location of the item at the scene; but they did not receive the initial examiner's notes or interpretation, medical findings, police theory, witness statements, etc. The research concluded that such a procedure was “both achievable and worthwhile” for bloodstain pattern analysis and potentially useful for other forensic disciplines [103].

Intelligence analysts have also been concerned about keeping biases away from the process of generating, evaluating, and not prematurely eliminating, multiple hypotheses [104]. In a debiasing technique called “Analysis of Competing Hypotheses”, analysts generate a set of hypotheses (mutually exclusive and preferably exhaustive) that might make sense of information. The analyst creates a matrix, with hypotheses in columns and pieces of relevant evidence in rows, and then evaluates the consistency of each hypothesis with each piece of evidence. This technique is supposed to help get rid of various biases including confirmation bias [105]. However, despite sounding like a good procedure, ACH has been subject to very little testing. Among the

potential problems with it is that it combines the evaluations in non-normative ways, including over-valuing uncertain inconsistent evidence relative to uncertain consistent evidence [106]. However, maintaining a list of considered hypotheses can be useful, as long as it is tucked away until needed, because constantly reviewing past ideas can block the generation of new ones [107].

5. Concluding thoughts

Humans can be great reasoners and terrible reasoners. We take top-down information (e.g., prior knowledge, expectations, motivation) and combine it with bottom-up information (e.g., what is out there in the world) and we infer what to think or how to act. And we typically get it right.

Forensic science demands discipline-specific analysis to reason in ways that are different from everyday reasoning. It asks them to ignore or set aside information that might lead them to an incorrect decision (e.g., police suspicions, gossip on the street) but also information that might help lead them to a correct decision for the wrong reasons (e.g., the results of other forensic analyses). It asks them not to engage in many of the types of reasoning described in Section 1 that help us accumulate, integrate, and assess information to make good decisions in everyday life. It does this so that other people in the criminal justice system, like police, lawyers, judges, and juries, can use the independent results of their work to reach properly-informed conclusions.

Optimizing analysts' decision making could include selecting people who can best operate under these unusual constraints (e.g., those high in need for cognition and low in need for cognitive closure [101,108]; see [44]; this issue); training in a particular manner (see [6]; this issue); and setting up an environment and standard operating procedures (see [48]; this issue, [109]; this issue) that facilitate the types of abnormal but essential reasoning that forensic science requires.

Funding Statement

This work was supported, in part, by the Forensic Technology Center of Excellence (Award 2016-MU-BX-K110) from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

References

- [1] J. Henrich, S.J. Heine, A. Norenzayan, Most people are not WEIRD, *Nature* 466 (7302) (2010) 29.
- [2] T. Busey, I. Dror, Special abilities and vulnerabilities in forensic expertise, in: A. McRoberts (Ed.), *The Fingerprint Sourcebook*, NLI Press, Washington, DC, 2011 (Ch. 15), <https://www.ojp.gov/pdffiles1/nij/225329.pdf>.
- [3] E. Balci, D. Dunning, See what you want to see: motivational influences on visual perception, *J. Pers. Soc. Psychol.* 91 (4) (2006) 612.
- [4] A. Raftopoulos, Is perception informationally encapsulated? The issue of the theory-ladenness of perception, *Cognit. Sci.* 25 (3) (2001) 423–451.
- [5] C.B. Mervis, E. Rosch, Categorization of natural objects, *Annu. Rev. Psychol.* 32 (1) (1981) 89–115.
- [6] Eldridge, H., Vanderkolk, J., & Stimac, J. (this issue). Learning from errors.
- [7] R.A. Searston, J.M. Tangen, K.W. Eva, Putting bias into context: the role of familiarity in identification, *Law Hum. Behav.* 40 (1) (2016) 50.
- [8] V.F. Holst, K. Pezdek, Scripts for typical crimes and their effects on memory for eyewitness testimony, *Appl. Cognit. Psychol.* 6 (1992) 573–587, <https://doi.org/10.1002/acp.2350060702>.
- [9] D. Davis, E.F. Loftus, Internal and external sources of misinformation in adult witness memory, in: *The Handbook of Eyewitness Psychology*, vol. 1, Psychology Press, 2017, pp. 195–238.
- [10] N. Pennington, R. Hastie, A cognitive theory of juror decision making: the story model, *Cardozo Law Rev.* 13 (1991) 519.
- [11] P. Thagard, Explanatory coherence, *Behavioral and Brain Sciences* 12 (1989) 435–502.
- [12] P. Thagard, Causal inference in legal decision making: explanatory coherence vs. Bayesian networks, *Appl. Artif. Intell.* 18 (3–4) (2004) 231–249.
- [13] S.J. Read, A. Marcus-Newhall, Explanatory coherence in social explanations: a parallel distributed processing account, *J. Pers. Soc. Psychol.* 65 (3) (1993) 429–477, <https://doi.org/10.1037/0022-3514.65.3.429>.
- [14] A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, *Science* 185 (4157) (1974) 1124–1131.
- [15] D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982.
- [16] B.W. Pelham, T.T. Sumarta, L. Myaskovsky, The easy path from many to much: the numerosity heuristic, *Cognit. Psychol.* 26 (2) (1994) 103–133.
- [17] S. Frederick, Cognitive reflection and decision making, *J. Econ. Perspect.* 19 (2005) 25–42.
- [18] B. Benson, Cognitive Bias Cheat Sheet, Simplified: Thinking Is Hard Because of 4 Universal Conundrums, Jan 8, 2017. Retrieved from, <https://medium.com/thinking-is-hard/4-conundrums-of-intelligence-2ab78d90740f>.
- [19] E.R. Smith, J. DeCoster, Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems, *Pers. Soc. Psychol. Rev.* 4 (2) (2000) 108–131.
- [20] D. Kahneman, *Thinking, Fast and Slow*, Macmillan, 2011.
- [21] R.E. Petty, J.T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, Springer Science & Business Media, 2012.
- [22] E. Pronin, Perception and misperception of bias in human judgment, *Trends Cognit. Sci.* 11 (1) (2007) 37–43.
- [23] J.T. Cacioppo, R.E. Petty, The need for cognition, *J. Pers. Soc. Psychol.* 42 (1982) 116–131.
- [24] G. Pennycook, J.A. Cheyne, D.J. Koehler, J.A. Fugelsang, Is the cognitive reflection test a measure of both reflection and intuition? *Behav. Res. Methods* 48 (2016) 341–348.
- [25] D. Kahneman, G. Klein, Conditions for intuitive expertise: a failure to disagree, *Am. Psychol.* 64 (6) (2009) 515.
- [26] J.M. Tangen, K.M. Kent, R.A. Searston, Collective intelligence in fingerprint analysis, *Cognitive research: principles and implications* 5 (2020) 1–7.
- [27] M.B. Thompson, J.M. Tangen, The nature of expertise in fingerprint matching: experts can do a lot with a little, *PLoS One* 9 (12) (2014), e114759. <https://journal.plosone.org/plosone/article?id=10.1371/journal.pone.0114759>.
- [28] J.M. Salerno, B.L. Bottoms, L.C. Peter-Hagene, Individual versus group decision making: jurors' reliance on central and peripheral information to evaluate expert testimony, *PLoS One* 12 (9) (2017), e0183580.
- [29] A. McCarthy Wilcox, N. NicDaied, Jurors' perceptions of forensic science expert witnesses: experience, qualifications, testimony style and credibility, *Forensic Sci. Int.* 291 (2018) 100–108.
- [30] G. Edmond, A. Towler, B. Grouns, G. Ribeiro, B. Found, D. White, K. Ballantyne, R.A. Searston, M.B. Thompson, J.M. Tangen, K. Martire, Thinking forensics: cognitive science for forensic practitioners, *Sci. Justice* 57 (2) (2017) 144–154.
- [31] B. Grouns, K.A. Martire, Human factors in forensic science: the cognitive mechanisms that underlie forensic feature-comparison expertise, *Forensic Sci. Int.: Synergy* 2 (2020) 148–153.
- [32] I.E. Dror, Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, *Anal. Chem.* 92 (12) (2020) 7998–8004.
- [33] A. McRoberts (Ed.), *The Fingerprint Sourcebook*, NLI, Washington, DC, 2011. <https://www.ojp.gov/pdffiles1/nij/225320.pdf>.
- [34] S.M. Kassin, I.E. Dror, J. Kukucka, The forensic confirmation bias: problems, perspectives, and proposed solutions, *Journal of applied research in memory and cognition* 2 (1) (2013) 42–52.
- [35] K. Lewin, *Principles of Topological Psychology*, McGraw Hill, 1936.
- [36] J.F. Rauthmann, R.A. Sherman, The situation of situation research: knowns and unknowns, *Curr. Dir. Psychol. Sci.* 29 (5) (2020) 473–480, <https://doi.org/10.1177/0963721420925546>.
- [37] G. Edmond, J.M. Tangen, R.A. Searston, I.E. Dror, Contextual bias and cross-contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals, *Law Probab. Risk* 14 (1) (2015) 1–25.
- [38] Z. Kunda, The case for motivated reasoning, *Psychol. Bull.* 108 (3) (1990) 480.
- [39] D.C. Murrin, M.T. Boccacini, L.A. Guarnera, K.A. Rufino, Are forensic experts biased by the side that retained them? *Psychol. Sci.* 24 (10) (2013) 1889–1897.
- [40] R.S. Nickerson, Confirmation bias: a ubiquitous phenomenon in many guises, *Rev. Gen. Psychol.* 2 (2) (1998) 175–220.
- [41] H. Eldridge, M. De Donno, C. Champod, Mind-set—How bias leads to errors in friction ridge comparisons, *Forensic Sci. Int.* 318 (2021) 110545.
- [42] J. Kukucka, S.M. Kassin, P.A. Zapf, I.E. Dror, Cognitive bias and blindness: a global survey of forensic science examiners, *Journal of Applied Research in Memory and Cognition* 6 (4) (2017) 452–459.
- [43] R.A. Searston, J.M. Tangen, The emergence of perceptual expertise with fingerprints over time, *Journal of Applied Research in Memory and Cognition* 6 (4) (2017) 442–451.
- [44] Spain, R. D., Hedge, J. W., Ohse, D., & White, A. (this issue). Personnel Selection and Assessment for the Forensic Sciences: an Overview of Methods and Research.
- [45] I. Dror, J. Melinek, J.L. Arden, J. Kukucka, S. Hawkins, J. Carter, D.S. Atherton, Cognitive bias in forensic pathology decisions, *J. Forensic Sci.* 66 (5) (2021) 1751–1757.
- [46] I.E. Dror, K.C. Scherr, L.A. Mohammed, C.L. MacLean, L. Cunningham, Biasability and reliability of expert forensic document examiners, *Forensic Sci. Int.* 318 (2021) 110610.
- [47] M.A. Almazrouei, I.E. Dror, R.M. Morgan, The forensic disclosure model: what should be disclosed to, and by, forensic experts? *International Journal of Law, Crime and Justice* 59 (2019) 100330.

- [48] Busey, T., Sudkamp, L., Taylor, M., & White, A. (this issue). Stressors in Forensic Organizations: Risks and Solutions.
- [49] W. Mannering, M. Vogelsang, T. Busey, F. Mannering, Are forensic scientists too risk averse? *J. For. Sci.* 66 (4) (2021) 1377–1400, <https://doi.org/10.1111/1556-4029.14700>.
- [50] T.D. Wilson, N. Brekke, Mental contamination and mental correction: unwanted influences on judgments and evaluations, *Psychol. Bull.* 116 (1) (1994) 117.
- [51] I.E. Dror, D. Charlton, Why experts make errors, *J. Forensic Ident.* 56 (4) (2006) 600.
- [52] I.E. Dror, D. Charlton, A.E. Péron, Contextual information renders experts vulnerable to making erroneous identifications, *Forensic Sci. Int.* 156 (1) (2006) 74–78.
- [53] S.M. Kassin, Why confessions trump innocence, *Am. Psychol.* 67 (6) (2012) 431–445.
- [54] S. Lewandowsky, U.K. Ecker, C.M. Seifert, N. Schwarz, J. Cook, Misinformation and its correction: continued influence and successful debiasing, *Psychol. Sci. Publ. Interest* 13 (3) (2012) 106–131.
- [55] K.A. Ranganath, B.A. Spellman, J.A. Joy-Gaba, Cognitive “category-based induction” research and social “persuasion” research are each about what makes arguments believable: a tale of two literatures, *Perspect. Psychol. Sci.* 5 (2010) 115–122, <https://doi.org/10.1177/1745691610361604>.
- [56] E.R. Tenney, R.J. MacCoun, B.A. Spellman, R. Hastie, Calibration trumps confidence as a basis for witness credibility, *Psychol. Sci.* 18 (1) (2007) 46–50.
- [57] M. Travers, L. Van Boven, C. Judd, The secrecy heuristic: inferring quality from secrecy in foreign policy contexts, *Polit. Psychol.* 35 (1) (2014) 97–111.
- [58] M.J.W. Wilson, B.A. Spellman, R. York, Beyond Instructions to Disregard: when Objections Backfire and Interruptions Distract, <http://dx.doi.org/10.2139/ssrn.2432527>, Saint Louis University, May 3, 2014. *Legal Studies Research Paper No. 2014-11*. Available at SSRN: <https://ssrn.com/abstract=2432527>.
- [59] J.B. Soll, K.L. Milkman, J.W. Payne, A user’s guide to debiasing, in: G. Keren, G. Wu (Eds.), *Wiley-Blackwell Handbook of Judgment and Decision Making*, 2015, ISBN 978-1118468395, pp. 924–951. Ch. 33.
- [60] K.A. Ericsson, R.R. Hoffman, A. Kozbelt, A.M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance*, second ed., Cambridge University Press, 2018.
- [61] W.C. Thompson, Determining the proper evidentiary basis for an expert opinion: what do experts need to know and when do they know too much? in: C. T. Robertson, A.S. Kesselheim (Eds.), *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law* Academic Press, 2016, pp. 143–150 (Chapter 9).
- [62] B.O. Gardner, S. Kelley, D.C. Murrice, K.N. Blaisdell, Do evidence submission forms expose latent print examiners to task-irrelevant information? *Forensic Sci. Int.* 297 (2019) 236–242.
- [63] B.O. Gardner, S. Kelley, D.C. Murrice, I.E. Dror, What do forensic analysts consider relevant to their decision making? *Sci. Justice* 59 (5) (2019) 516–523. <https://www.sciencedirect.com/science/article/pii/S1355030618302867?via%3DIihub>.
- [64] N. Stebbly, H.M. Hosch, S.E. Culhane, A. McWethy, The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: a meta-analysis, *Law Hum. Behav.* 30 (4) (2006) 469–492.
- [65] J. Kukucka, I.E. Dror, M. Yu, L. Hall, R.M. Morgan, The impact of evidence lineups on fingerprint expert decisions, *Appl. Cognit. Psychol.* 34 (5) (2020) 1143–1153.
- [66] A. Quigley-McBride, G.L. Wells, Fillers can help control for contextual bias in forensic comparison tasks, *Law Hum. Behav.* 42 (4) (2018) 295.
- [67] K.N. Ballantyne, G. Edmond, B. Found, Peer review in forensic science, *Forensic Sci. Int.* 277 (2017) 66–76.
- [68] E.J. Mattijssen, C.L. Witteman, C.E. Berger, R.D. Stoel, Cognitive biases in the peer review of bullet and cartridge case comparison casework: a field study, *Sci. Justice* 60 (4) (2020) 337–346.
- [69] Innocence Project Website, DNA exonerations in the United States. <https://www.innocenceproject.org/dna-exonerations-in-the-united-states/>, retrieved Aug 3, 2020.
- [70] L.J. Curley, J. Munro, M. Lages, An inconvenient truth: more rigorous and ecologically valid research is needed to properly understand cognitive bias in forensic decisions, *Forensic Sci. Int.* 2 (2020) 107.
- [71] S.V. Stevenage, A. Bennett, A biased opinion: demonstration of cognitive bias on a fingerprint matching task through knowledge of DNA test results, *Forensic Sci. Int.* 276 (2017) 93–106.
- [72] J. Kukucka, People who live in ivory towers shouldn’t throw stones: a refutation of Curley et al, *Forensic Sci. Int.: Synergy* 2 (2020) 110.
- [73] W.R. Oliver, Comment on Kukucka, Kassin, Zapf, and Dror (2017), “Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners”, 2018.
- [74] I.E. Dror, J. Kukucka, S.M. Kassin, P.A. Zapf, When expert decision making goes wrong: consensus, bias, the role of experts, and accuracy, *Journal of Applied Research in Memory and Cognition* 7 (1) (2018) 162–163, <https://doi.org/10.1016/j.jarmac.2018.01.007>.
- [75] N.J. Roesse, K.D. Vohs, Hindsight bias, *Perspect. Psychol. Sci.* 7 (5) (2012) 411–426.
- [76] J. Valley, Metro reviewing DNA cases after error led to wrongful conviction. Las Vegas Sun, Retrieved from: <https://lasvegassun.com/news/2011/jul/07/dn-a-lab-switch-led-wrongful-conviction-man-who-ser/>, July 7, 2011.
- [77] Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination And Human Factors: Improving The Practice Through a Systems Approach*, U.S. Department of Commerce, National Institute of Standards and Technology, 2012, <https://doi.org/10.6028/NIST.IR.7842>.
- [78] President’s Council of Advisors on Science and Technology [PCAST], *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, September, 2016.
- [79] A.L. Alter, D.M. Oppenheimer, Uniting the tribes of fluency to form a metacognitive nation, *Pers. Soc. Psychol. Rev.* 13 (3) (2009) 219–235, <https://doi.org/10.1177/1088868309341564>.
- [80] W.C. Thompson, Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation, *Law Probab. Risk* 8 (2009) 257–276.
- [81] J. Kukucka, S.M. Kassin, Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias, *Law Hum. Behav.* 38 (3) (2014) 256–270.
- [82] *United States v. Trenkler*, 61 F.3d 45 (1st Cir. 1995).
- [83] I.E. Dror, C. Champod, G. Langenburg, D. Charlton, H. Hunt, R. Rosenthal, Cognitive issues in fingerprint analysis: inter-and intra-expert consistency and the effect of a ‘target’ comparison, *Forensic Sci. Int.* 208 (1–3) (2011) 10–17.
- [84] H. Swofford, S. Steffan, G. Warner, C. Bridge, J. Salyards, Inter- and intra-examiner variation in the detection of friction ridge skin minutiae, *J. Forensic Ident.* 63 (5) (2013) 553–570.
- [85] J.R. Vanderkolk, Examination processes, in: A. McRoberts (Ed.), *The Fingerprint Sourcebook*, 2011. Ch. 9. NJL, <https://www.ncjrs.gov/pdffiles1/nij/225329.pdf>.
- [86] D.L. Medlin, R.L. Goldstone, D. Gentner, Respects for similarity, *Psychol. Rev.* 100 (2) (1993) 254–278.
- [87] A. Tversky, Features of similarity, *Psychol. Rev.* 84 (4) (1977) 327–352.
- [88] K.A. Carlson, M.G. Meloy, J.E. Russo, Leader-driven primacy: using attribute order to affect consumer choice, *J. Consum. Res.* 32 (4) (2006) 513–518.
- [89] D.R. Ashbaugh, *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*, CRC Press, Boca Raton, FL, 1999.
- [90] I.E. Dror, W.C. Thompson, C.A. Meissner, I. Kornfield, D. Krane, M. Saks, M. Risinger, Letter to the editor-context management toolbox: a linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making, *J. Forensic Sci.* 60 (4) (2015) 1111–1112.
- [91] G. Langenburg, Addressing potential observer effects in forensic science: a perspective from a forensic scientist who uses linear sequential unmasking techniques, *Aust. J. Forensic Sci.* 49 (5) (2017) 548–563.
- [92] AAAS, *Forensic Science Assessments: A Quality And Gap Analysis – Fire Investigation* (Report Prepared by Jose Almirall, Hal Arkes, John Lentini, Fred Mowrer, & Janusz Pawliszyn, June 2017, <https://doi.org/10.1126/srhl.aag2872>.
- [93] P. Dehghani-Tafti, P. Bieber, Folklore and forensics: the challenges of arson investigation and innocence claims, *W. Va. Law Rev.* 119 (2) (2016) 549–620.
- [94] N.K. Osborne, M.C. Taylor, M. Healey, R. Zajac, Bloodstain pattern classification: accuracy, effect of contextual information and the role of analyst characteristics, *Sci. Justice* 56 (2) (2016) 123–128.
- [95] N.K. Osborne, M.C. Taylor, R. Zajac, Exploring the role of contextual information in bloodstain pattern analysis: a qualitative approach, *Forensic Sci. Int.* 260 (2016) 1–8.
- [96] P. Bieber, *Fire Investigation and Cognitive Bias*, *Wiley Encyclopedia of Forensic Science*, Wiley Online Library, Dec 11, 2014. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470061589.fsa1119>.
- [97] W.C. Thompson, N. Scurich, When does absence of evidence constitute evidence of absence? *Forensic Sci. Int.* 291 (2018) e18–e19.
- [98] P. Cherubini, E. Castelvechio, A.M. Cherubini, Generation of hypotheses in Watson’s 2–4–6 task: an information theory approach, *The Quarterly Journal of Experimental Psychology Section A* 58 (2) (2005) 309–332.
- [99] D. Kuhn, M. Weinstock, R. Flaton, How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task, *Psychol. Sci.* 5 (5) (1994) 289–296.
- [100] P.E. Tetlock, *Expert Political Judgment: How Good Is it? How Can We Know?*, 2005/2017. Princeton.
- [101] D.M. Webster, A.W. Kruglanski, Cognitive and social consequences of the need for cognitive closure, *Eur. Rev. Soc. Psychol.* 8 (1) (1997) 133–173.
- [102] A.C. Doyle, *The Sign of the Four*, 1890.
- [103] N.K. Osborne, M.C. Taylor, Contextual information management: an example of independent-checking in the review of laboratory-based bloodstain pattern analysis, *Sci. Justice* 58 (3) (2018) 226–231.
- [104] B.A. Spellman, Individual reasoning, in: B. Fischhoff, C. Chauvin (Eds.), *Intelligence Analysis: Behavioral and Social Scientific Foundations*, National Academies Press, Washington, DC, 2011, pp. 117–141.
- [105] R.H. Heuer Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence Analysis, 1999.
- [106] D.R. Mandel, C.W. Karvetski, M.K. Dhami, Boosting intelligence analysts’ judgment accuracy: what works, what fails? *Judgment and Decision Making* 13 (6) (2018) 607–621.
- [107] S.M. Smith, T.B. Ward, J.S. Schumacher, Constraining effects of examples in a creative generation task, *Mem. Cognit.* 21 (6) (1993) 837–845.
- [108] J.T. Cacioppo, R.E. Petty, J.A. Feinstein, W.B.G. Jarvis, Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition, *Psychol. Bull.* 119 (2) (1996) 197.
- [109] Carlson, L., Kennedy, J., Zeller, K., & Busey, T. (this issue). Communication during a Forensic Investigation: the Pebbles on a Scale Conceptual Model.