UNIVERSITY OF PADOVA

Department of General Psychology

Master Degree in Cognitive Neuroscience and Clinical Neuropsychology

Final dissertation

Cognitive Bias in Neuroimaging: A Study on How Information Affects Neuroradiological Reports

Supervisor

Professor Cristina Scarpazza

Candidate: Janet Louisa Wijaya

Student ID number: 2002399

Academic Year 2021-2022

1. Introduction

Real-life situations require making a series of decisions (Prezenski et al., 2017). As humans, we make many judgments and decisions of different complexity and importance every day, depending on previous feedback from a potentially changing environment (Glöckner & Witteman, 2009; Prezenski et al., 2017). Unlike static problems, such as arithmetic problems, the constant change in environment leads to dynamic situations where any complication in the environment could potentially cause a shift in a previously set decision (Dörner & Funke, 2017; Edwards, 2016). Take an everyday task such as going to the store as an example. Before leaving the house, we decide which store to go to depending on what we want to buy and previous store experiences, such as item availability or service. We also consider other variables that may affect the journey, such as the weather or the traffic situation at that moment. The described scenario illustrates an example of the use of complex cognition, which is all mental processes that individuals use for deriving new information out of given information, with the intention to solve problems, make a decision, and plan actions (Knauff & Wolf, 2010). The mental activities such as thinking, reasoning, problem-solving, and decision-making that makes up this complex cognition, typically rely on simple cognition, which is the combination and interaction of more elementary processes such as learning, memory, emotion, and perception (Knauff & Wolf, 2010; Sternberg & Ben-Zeev, 2001).

1.1 Complex Cognition and Simple Cognition

Does this mean that complex cognition is simply the continuum of simple cognition? On the one hand, simple cognition is involved in complex cognition. On the other hand, complex cognition is more than the summation of the processes of simple cognition. Following the Gestalt principle, complex cognition is a system, where its constituent parts are interrelated with each other and with the system (Funke, 2010). In the interconnected system, each part and subpart have functions that contribute to the wholeness of the system (Wertheimer, 1985, 2010).

Take visual processing, for example. Model of visual processing states that visual information is processed in 2 ways: bottom-up (data-driven) and top-down (context-driven) (Pylyshyn, 1999). Bottom-up refers to processing that is driven purely by sensory input. In contrast, top-down refers to processing driven by stored information (i.e., context) and expectations (Ganis & Kosslyn, 2007). The interaction between bottom-up and top-down processing is important to produce reliable information processing, especially if there are ambiguities in the stimulus. Consider the example of the Mooney face illusion. The Mooney face illusion is a human face stimulus with the addition of round patches of black and white (Schwiedrzik et al., 2018). The patches cause the content of the stimulus to become unclear, creating the possibility of someone interpreting the stimulus as a face or as a saxophone player. Like other illusions, such as the young-old-woman illusion, the Mooney face shows that what we perceive at first glance depends on our top-down processing (Carbon, 2014). If we think of a woman before viewing the picture, the chance that we interpret the picture as a woman increases. Instead, if we think of a musician before viewing the picture, the chance that we interpret the picture as a saxophone player increases.

Figure 1 Mooney C. (1950). Mooney face picture. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:MooneyFace.png

Simple cognition here represents bottom-up processing, where the perception of the stimulus solely indicates that our visual system can see a black-and-white image. However, it does not answer why we can switch between perceiving a face or a saxophone player. When we shift to complex cognition, we embed the more rudimentary process, the perception of the black-and-white image, into an interconnected system that aims to identify the picture (Funke, 2010). The image is now considered along with another context, such as prior expectations and other contextual information (i.e., was there music playing nearby when we were looking at the picture), which, modulates the interpretation of the image into a face or a saxophone player.

When we shift from simple to complex cognition, we shift the hierarchy level of explanation (Funke, 2010). As we move up the cognitive process, we consider the black-and-white image more than just signal detection. Information used to interpret the Mooney face comes from perception and the accompanying contexts (Funke, 2010).

1.2 Complex Problems

A problem is conceptualized as a composition of a given state, the desired goal, and the obstacle between the two states (Mayer, 1992). A problem then is said to become complex when finding the solution demands a series of operations which can be characterized as follows (Dörner et al., 1983): Elements relevant to the solution process are significant (complexity), highly interconnected (connectivity), and dynamically changing over time (dynamics). Furthermore, neither structure nor dynamics are disclosed (intransparency). Lastly, the goal structure is not straightforward (Funke, 2010).

Information gathered through perception alone is insufficient in resolving a complex problem. While our senses constantly work to resolve uncertain information about the current state of the world, it is not enough. For example, our vision works quite well in the daytime, but it gradually loses accuracy as lights dim, and we are blind to what is happening behind our backs. Furthermore, when looking at ambiguous visual stimuli, the observer experiences frequent spontaneous transitions between two competing percepts even when the physical stimulation remains unchanged (Sterzer & Kleinschmidt, 2007). The nature of the complex problem requires people to engage in complex cognition, as reliance on simple cognition alone leads to information deficiency or uncertainty (Ayyub & Klir, 2006; Funke, 2010).

Uncertainty is a common experience (Hammond, 2000) and an intrinsic part of neural computation and information processing, whether for sensory processing, motor control, or cognitive reasoning. For instance, it is impossible to determine the age of a person based on a photo, but it is possible to reasonably guess and even estimate the uncertainty associated with that guess (Gershman & Beck, 2017; Pouget et al., 2013). Uncertainty adds a particular layer of complexity to information processing, and intuitively, we may think that a complex judgment task requires complex solutions. However, that may not be the case. Humans mostly do not use complex cognitive strategies to make reasonable inferences, estimations, and other judgments (Marewski et al., 2010).

Imagine a baseball player who wants to catch a ball high up in the air in his direction. How would the player solve the complex task of catching the ball? Following the idea of complex cognition, the player should solve multiple equations to predict the ball's trajectory while including parameters such as speed, wind current, and the spin of the ball. Then the player could calculate precisely where the ball would land and run to that location (Davis, 2017; Marewski et al., 2010).

However, solving all those calculations would seem unlikely due to time and biological constraints (Hoc & Amalberti, 2016; Noppeney, 2021). In fact, to catch the ball, one simply must fixate it, start running, try to keep the angle constant (Marewski et al., 2010). By fixating on the ball, the player would be able to be exactly in the desired position, without performing any of the trajectory computations. Complex judgment tasks sometimes do not need complex cognitive strategies to be solved successfully (Marewski et al., 2010).

1.3 The Difficult Task of Inference and Heuristics

Judgment is the process by which individuals consider and evaluate evidence and estimate the likelihood of occurrence of different outcomes (Blanchette & Richards, 2009). As humans make judgments, we not only have access to external information provided by the environment but also receive meta-cognitive cues from our internal mental processes. One example of meta-cognitive cues we use is fluency — the subjective feeling of ease or difficulty that arises when we process information (Oppenheimer, 2008). The limitation of information processing due to biological constraints of the brain drives the brain to pay attention to fluency when reasoning. Rather than being a cognitive operation, fluency is the feeling of ease that can accompany a cognitive operation. It can theoretically be generated by nearly any form of thinking, and in turn, it can serve as a cue for judgments in any situation (Oppenheimer, 2008). For example. easily retrieved instances are judged as frequent (Tversky et al., 1973; Unkelbach, 2006), and fluently perceived names are judged as famous (Jacoby et al., 1989; Unkelbach, 2006). People draw on metacognitive experiences to assess their knowledge and various task characteristics. These assessments then inform the choice of processing strategies that are likely to influence judgment, such as the use of heuristics. (Schwarz, 2010).

In order to arrive at quick solutions to perceptual problems, our brain turns to heuristics. Heuristics are strategies that reduce the load of information processing, that is, by guiding information search and modifying the problem's representation, therefore streamlining event probability calculation and judgmental operation (Goldstein & Gigerenzer, 2002; Tversky & Kahneman, 1974). For example, consider this description of an individual: "John is very quiet and polite, and he tends to keep his composure when he faces stressful situations. He is very detail-oriented, tidy, and appreciates order and structure in his life. In his free time, he likes to venture to new places". Now, suppose someone were to ask to assign an occupation for John from a list of occupations (i.e., farmer, salesman, pilot, librarian, or physician). In that case, we are likelier to assign a high probability towards the pilot option. We choose to assign this probability not only due to his description but also due to a heuristic called representativeness.

1.4 Representativeness Heuristic and Errors

The representativeness heuristics reflects the probabilities examined through the problems, such as to which degree object A belongs in the same class as object B. When A highly resembles B, the probability that A originates from B is then deemed to be high. On the contrary, if A does not resemble B, the probability that A originates from B is then deemed to be low (Tversky & Kahneman, 1974). For example, if the prototype of a pilot is a calm, responsible, and adventurous, then John resembles the description quite well. When using the representativeness heuristic, the probability that John is a librarian is assessed by how similar he is to the stereotype of a librarian. Heuristics are useful to simplify information processing tasks. However, solutions achieved via heuristics do not always mean the solution is optimal. Heuristics often lead us to results that are "good enough" but can sometimes lead to severe and systematic errors (Tversky & Kahneman, 1974).

Returning to the case of John's occupation, the fact that the population of physicians is way bigger than the population of pilots should also be considered, as John could very well be a physician who likes to travel in his free time, for example. However, considerations of the population of pilots and physicians, or base-rate frequency, do not affect the degree of similarity of John to the stereotype of a pilot. Prior probabilities tend to be ignored if we rely on representativeness heuristics (Tversky & Kahneman, 1974). Thus, being overly reliant on heuristics leads to cognitive biases — systematic selectivity in the information processing system that favors one type or source of information over another (MacLeod & Mathews, 2012).

1.5 Cognitive Biases and Decision-Making in Experts

We are constantly making decisions that vary in importance, from choosing what fruit to buy in the market, to deciding which treatment option to take for an illness or whether the defendants are guilty of their charged crimes. Throughout this decision-making process, cognitive biases, which are systematic unconscious distortions in thinking, are pervasive and affect our judgments (Blanco, 2017). Suppose we return to the John example. After reviewing the description, we concluded that John was most likely a pilot. However, if John was actually a physician, our judgmental error could be attributed to the existence of cognitive biases. Another common bias could be the familiarity/availability bias— the estimation of the likelihood of an event being affected by how easy it is to recall similar events. Confirmation bias occurs when people search for evidence that confirms their idea rather than the opposite.

Finally, overconfidence bias occurs when people overestimate their prediction accuracy or their performance as more remarkable than it actually is (Ellis, 2018).

When these cognitive biases influence decision-making, it could result in a negative outcome. Now consider the role of eyewitnesses in a law enforcement setting. Eyewitnesses play an important role in apprehending criminals and all the process that follows (i.e., providing testimonies that provide information for the jury) until a verdict is reached (Brigham et al., 1982; Wells et al., 2003). However, many scientists, legal scholars, and law enforcement personnel have suggested that, perhaps, eyewitnesses are not as accurate as we once thought (Brigham et al., 1982; Wixted et al., 2018). Studies do show that the performance of eyewitnesses is, frankly, quite low (Wise et al., 2014). As time passes, witnesses tend to make more and more recall errors and recognition errors (Pansky et al., 2005), approaching chance performance within just one day (Brigham et al., 1982). What is more, they tend to underestimate the descriptions of characteristics (such as weight, height, and age) while they usually provide a description that resembles the population's average instead of the characteristics of the suspect itself (Brigham et al., 1982; Buckhout et al., 1975). Other variables such as sub-optimal memory conditions (i.e., when the person is under high stress or when distractions are present) and interview conditions could also lower the reliability of the eyewitness testimonies (Castelli et al., 2006; Deffenbacher, 2008; Wixted et al., 2018). For example, exposure to mugshots of a suspect during the interview leads to mugshot-induced bias, which increases the likelihood that the witness chooses that suspect from a lineup later (Magnussen et al., 2010).

Experts tend to overlook to what extent the presence of errors, cognitive biases, or other factors, affect eyewitness testimonies. For example, several surveys found that when compared to eyewitness experts, knowledge about factors that affect the reliability of eyewitnesses has not reached a common sense level among professionals in the judiciary systems in Europe and USA (Benton et al., 2006; Granhag et al., 2005; Magnussen et al., 2009, 2010; Wise & Safer, 2004), but law professionals are more knowledgeable about factors that affect eyewitness reliability than potential jurors (Benton et al., 2006). Also, no correlation has been found between the number of times one serves on jury duty and the knowledge about factors that can affect eyewitness reliability (i.e., weapon focus, minor details, accuracy impairments due to stress, forgetting curve) (Magnussen et al., 2010). Hence, the field of forensic science needed a more reliable line of evidence, as eyewitnesses, and many other lines of evidence (i.e., anecdotal, analogical, characteristic), are inherently weak in reliability and even problematic (Wells et al., 2003).

1.6 Forensic Science and Its Potential Problem

Forensic science — the application of science to aid the criminal justice system processes— has become an integral part of the justice system. Forensic science provides information that aids in investigating and prosecuting crime through scientific examination and analysis of the physical evidence. It could also provide insight into the characteristics and behavior of the criminal (Julian et al., 2011; Saferstein, 2018). Indeed, over the last 40 years, considerable advancements have pushed forensic science to act as an investigative and intelligence tool for police officers as they pursue criminal cases (Peterson et al., 2013). Advances in methods of identifying biological materials (e.g., fingerprinting or DNA profiling) found in crime scenes have changed the way police conduct criminal investigations, particularly in increasing the possibility of linking a suspect to the crime scene while also removing the innocent parties from the ongoing investigation (Julian et al., 2011).

The contribution of evidence obtained from forensic science often overpowers the other lines of evidence that are more susceptible to inaccuracies either due to the inherent unreliability of the evidence itself or due to the users' ignorance. Hence, it is not surprising that forensic science has an increasing and often decisive authority in the judicial system (Dror, 2015). This increase comes from the assumption that forensic science provides material for experts, who can presumably offer objective and impartial scientific input at trial. However, it would be unwise to blindly believe that experts interpret the evidence as infallible (Nakhaeizadeh et al., 2015). A notable example came from a case from Madrid, Spain. On March 11, 2004, bomb explosions on four commuter trains killed 191 people and wounded 1800 others, setting a full-scale international investigation into motion. The U.S. Federal Bureau of Investigation (FBI) positively identified Brandon Mayfield as the culprit based on fingerprint evidence. Following the standard protocol, other independent examiners also concluded that Mayfield was the culprit. It was important to note that the investigation, arrest, and trial were subsequently made after the 9-11 incident. However, soon the Spanish authorities matched the prints to the actual Madrid bomber, Ohnane Daoud. The culprit mismatch led to the FBI being investigated, and a report by the Office of the Inspector General (OIG, 2006) concluded that "confirmation bias" was listed as a contributing factor to the erroneous identification (Kassin et al., 2013).

Fingerprints and several other forensic science testimonies have also been used to convict innocent individuals wrongfully. In cases where trial transcripts or reliable forensic science data were available for review, 38% contained incorrect serology testimony. In addition, 22% involved hair comparisons; 3% involved bite mark comparisons; and 2% involved fingerprint comparisons (Kassin et al., 2013).

1.7 Human Factors in Forensic Science

The issue surrounding the fallibility of forensic science lies behind its human examiners (Dror, 2015). Ever since the emergence of forensic science about 100 years ago, there has been systematic neglect of human examiners, who play a crucial role in the field as they are the primary agent of analysis (Dror, 2015). Even in other lines of evidence that rely more on "objective" quantification and instrumentation, like fingerprint analysis, human intervention influences the whole process. Human influence affects most stages of the forensic work, from the initial stages of sampling, determining what noise is and what should be used as input, to the final stages of communicating the result; it also plays a role in issues such as biases in decision-making (Kassin et al., 2013). The goal of forensic science is to create forensic decisions. However, because scientists are humans, the element of subjectivity is involved in this endeavor. Because of this, the cognitive processes in question are vulnerable to various biases that come from external factors, such as contextual information (Stoel et al., 2016). With the potential of biases arising from contextual information, forensic examiners must focus on the relevant scientific data, isolating and blinding themselves from information that they do not need and that can bias their forensic work (Dror, 2015).

For instance, a study was done on forensic pathologists who serve as coroners to examine biases regarding decisions about the cause of death. The primary data provided are death certificates, which are documents made by legal practitioners that contain information regarding someone's death. The study (Dror et al., 2021) presents 133 forensic pathologists with identical medical information about a child's death but randomly assigns them with different medically irrelevant contextual information and asks them to determine the manner of death (categorizing death as accident or homicide). It is important to note that non-medical contextual information may not only be considered when determining the manner of death but must be considered. By its very nature of being circumstantial information, the manner of death depends on the investigation, which may reveal the circumstances (and, as necessary, the broader background, e.g., medical history) surrounding the death. The specific contextual information chosen in the experimental study, race and caretaker identity, was purposely (and adequately) designed to be always irrelevant for determining the manner of death. Such a study can reveal how cognitive biases impact forensic pathologists' decisions and conclusions. Results of the study reveal that Black children were more often judged to be victims of homicides than White children, who were more often judged to be victims of accidents.

Similarly, children whose caretaker was a grandmother were more likely to be considered to have died because of an incident. In contrast, children whose caretaker was the child's mother's boyfriend are more likely to have been murdered. Information, such as the child's race and the caretaker's identity, is irrelevant to the decision maker. The pathological decision, which is the manner of death, should only be driven by medically relevant information and supplemented by less medically relevant information when needed and justified. For example, the child's race could be a potential source of bias that affects the base rate. Suppose a forensic pathologist encountered plenty of cases where a Black child has died due to a homicide rather than an accident. In that case, the pathologists will develop an a priori expectation that Black children are a victim of homicide, regardless of the accuracy. Hence, to see that this information plays a factor in decision-making shows the existence of cognitive biases. It shows that even highly trained professional scientists can be biased in their decisions (Dror et al., 2021).

Cognitive biases could impact how data are perceived, how data are interpreted, and eventually, how conclusions are reached (Dror, 2020). Because of this, it becomes important to present and dispel any commonly held fallacies where the biases stem from, as it would be impossible to advance if we minimize or dismiss the existence of these biases. Therefore, the first step is acknowledging their existence (Dror, 2020; Edmond et al., 2017).

1.8 Fallacies

One relevant fallacy is expert immunity, which is the incorrect belief that experts are impartial and immune to biases. However, the truth is, as we can see from the study above, no one, not even experts, is immune to biases. Experts are more susceptible to certain biases, and the susceptibility results from reaching the expertise level. For example, through experience and training, experts are more prone to engage in selective attention and use chunking and schemas (Dror, 2020). Experts also tend to rely on heuristics and expectations based on their experience (Islam et al., 2014; Perez, 2014), all while still using various top-down cognitive processes to form their a priori assumptions and expectations (Dror, 2020). Nevertheless, all the above techniques do assist experts in making quick decisions.

On the other hand, experts are now more susceptible to biases that could lead them to the wrong conclusion (Dror, 2020). Erroneous conclusions point not only to the susceptibility of experts towards cognitive biases in general but also to the fact that more experienced experts, who tend to be more confident, could potentially perform worse than novices. The worsening of experts' performance has also been demonstrated in other fields, for example, in environmental ecology, where it was shown that novice experts outperform experienced experts in terms of the correct identification of collected data (Dror, 2020).

Another example of a fallacy is the bias blind spot; humans, including experts, are often unaware of their biases, leading to the false belief that they are not biased (Dror, 2020; Kukucka et al., 2017). While it is relatively easier to point out biases in others, biases are harder to identify when they are within us, and sometimes we are even blind to the existence of the biases themselves. For example, a research found that 70% of forensic scientists now acknowledge cognitive biases as a cause of concern in the forensic field. However, only 52% think it is a concern for their specific fields, and only 25% think it is a concern for them personally, unknowingly reflecting the bias blind spot fallacy (Dror, 2020).

The last example of a fallacy is the illusion of control (Meissner & Wulf, 2017); when experts are told about their biases, they believe that sheer willpower alone is enough to overcome them. However, countering the effect of the biases requires taking specific steps. Hence, willpower alone is not enough to combat and deal with the various manifestation of these biases (Dror, 2020). Trying to deal with biases under the illusion of control could increase the chance of encountering the biases themselves. The increase in chance is explained through the ironic processing of mental control theory, in which the desire to control a mental state can yield the opposite result from what was intended (Dugdale & Eklund, 2002; Wegner, 1994). To illustrate the theory, it is like a judge instructing jurors to disregard certain evidence. By doing so, the judge makes the jurors notice this evidence even more.

1.9 Origins of Cognitive Biases: Framework by Dror (2020)

Cognitive biases can arise from a variety of sources, which are categorized into three groups. Category A relates to the specific case — something about this case causes bias in how data are perceived, analyzed, and interpreted. Other sources of cognitive biases, Category B, have nothing to do with the specific case. However, they arise from factors relating to the specific person doing the work — something about him/her (e.g., his/her experience, personality, working environment, or motivation) causes the biases. Finally, other sources of cognitive biases, Category C, arise from human nature, the very cognitive architecture of the human brain that we all share, regardless of the specific case or the person doing the analysis (Dror, 2020).

Starting from Category A, the data is the first source of cognitive biases within the factors that relate to the specific analysis. How can data cause biases? Well, it depends on the data. Some data, such as finger marks, do not, per se, cause biases, as they convey no information beyond the friction ridges impressions. However, with other types of data (such as in the analysis of voice, handwriting, blood spatter, bitemarks, and psychiatric symptoms), the data can contain potentially biasing information. The second source comes from the reference materials used as the comparative material for the evidence itself. Perception and interpretation of the data are affected by their comparative counterpart, which applies to various data requiring comparison (i.e., fingerprints, blood samples, DNA profiles). For example, if the suspect's DNA profile is used as the reference material, interpretation of the biological material obtained from the crime scene will be biased and tend to shift to fit the reference better. Due to this, decisions are driven by the suspect's profile instead of a data-driven interpretation, where the evidence is interpreted based on the data it contains. Therefore, if the investigation goes from the direction of the evidence to the suspect, the reference materials instead cause the examiners to reason from the suspect to the evidence, creating circular reasoning and biasing the interpretation of the data. The third source comes from contextual information. Experts are often exposed to irrelevant information. In the forensic domain, for example, such information may be that the suspect confessed to the crime, that eyewitnesses and other lines of evidence have identified them, or that the suspect has a criminal record. Knowing the suspect's name suggest a specific race, evoking biases and stereotypes. These all cause expectations that can impact not only the interpretation of the results obtained from the analysis but also the analysis itself because the expectations impact the detection of what goes into the analysis and testing strategies. This source of bias is not derived from a target generated by the reference materials but from contextual information that can be irrelevant to the task carried out by the analyst. Contextual expectations impact not only data collection and testing strategies but also the interpretation and conclusions of the analysis. Cognitive biases arise when task-irrelevant context causes some aspect of analysis to be over-weighted, under-weighted, or neglected (e.g., not perceived, determined to be noise, an anomaly, or an outlier). The worsening of performance due to cognitive biases does not only happen in subjective judgments but can also bias established procedures and criteria for accepting evidence and proper judgment. The problem with task-irrelevant contextual information is that it can cause many kinds of biases that impact analysis in many ways. Another impact of biases can be overlooking or underweighting the absence of data, not properly confirming results, or not considering alternatives. It is important to emphasize that contextual irrelevant information biases scientists and experts, and it can do so at an unconscious level−they may not be aware of the impact. The expectation created by contextual information also biases what and how information is represented and processed in the brain. These biases impact experts and cannot be adequately controlled by mere willpower (Dror, 2020).

Moving on to Category B, a possible source of biases comes from the base rate. In the forensic field, experts create base rate probability through experience obtained from previous cases. Although an important asset, it also impacts the interpretation of the data due to building certain expectations toward new cases derived from experience instead of the specific case currently at hand. Sampling and analysis could potentially be affected by these expected base rates, which are unrelated to the current case. Other stages, such as detection and verification of evidence (i.e., marks, signals, items), could also be affected. For example, base rate bias for low target prevalence shows that if in past experiences it was rare or uncommon to find, observers are more likely to miss it in the future. The low base rate, of course, biases the search done by the experts, where rare and uncommon items tend to be neglected even if they are present. Base rate bias derives from expectations generated from past similar cases. The issue is that this case is biased because its analysis is based on other cases. The crux of the bias is that perception and decisions are not based on the case itself. This bias is even more potent when the similarity to past cases is superficial and more in the eye of the beholder. The fifth source is organizational factors. Organizational factors that can cause biases are many and varied and have been well documented in various domains. When it comes to DNA and other forensic evidence, where analysis and work are often conducted within the adversarial legal system, cognitive biases may emerge from an allegiance effect and myside bias. Indeed, a study showed that when forensic experts are presented with identical data, they reach conclusions biased toward the side that retained them — an adversarial allegiance and myside bias. These are implicit biases, not explicit partiality, when one side is openly favored over the other. The impact of organizational factors applies to every laboratory — they work within various contexts, structures, and frameworks that can bias their work. For example, laboratories have a clear hierarchy and dependencies. If a senior person “signs off” on reports or analyses, there can be the danger of “writing what that person wants to read” and a lack of challenge in their scientific decisions. Thus, science is muddled with managerial authority and other organizational pressures. Other organizational factors relate to time pressure, expectations to reach certain results, stress, budget controls, pressure to obtain publications and other targets, and various organizational factors that can impact the work carried out in laboratories and other professional settings. The sixth comes from education and training. Education and training play an important role in how work is conducted. For example, forensic examiners see their role more as supporting the police rather than as scientists. When approaching a case, training and education may instill the pursuit of a single hypothesis vs. examining multiple hypotheses, considering alternative hypotheses (including scenarios proposed by the opposing side), conducting a differential diagnosis, and considering categorical decisions (such as "match the opposing" and "non-match", often used in fingerprinting and firearms) vs. using statistics and other methods to determine the strength of the evidence. The seventh source comes from personal factors. Many personal factors impact biases and decision-making. These include motivation, personal ideology, and beliefs. Furthermore, some people are risk-takers and others are risk-averse, and people also vary their tolerance to ambiguity. Other individual differences between people can bias results. These factors are minimized in areas where there is more objective quantification and instrumentation. However, in areas where the human examiner has a greater role in deciding how to collect, sample, and interpret the data and where there is subjectivity in evaluating the data and conclusions, such personal factors play a greater role in how work is carried out. Other personal factors that can cause bias in decisions include the need for closure that can result in hasty decisions or opting to reach inconclusive decisions, how people respond to stress and fatigue, personality, and a whole range of personal factors that can impact expert decision making (Dror, 2020).

Lastly, from Category C, the source comes from the human and cognitive factors of the brain. The workings of our brain create architectural and capacity constraints that do not allow it to process all the incoming information. The brain, therefore, engages in various processes (mainly known as "top-down") to make sense of the world and data around us. The human mind is not a camera. The active feature of human cognition means that we do not see the world "as it is." Beyond many cognitive processes and how the human brain is wired, which can cause biases, there are biasing effects related to social interaction, in-group and availability biases, processing fluency, and other biasing influences that impact all of us (Dror, 2020). In ambiguous decision-making criteria, the potential for biases is enhanced, especially through the sources explored above. Although some referral questions in forensic psychology are well defined (e.g., diagnostic evaluations, competency to stand trial) and have decision aids that can help the evaluator more easily answer the question (e.g., testing materials with explicit decision rules or actuarial formulas), the nature of some referral questions are not so straightforward. In these evaluations, the potential for examiner bias is greater (Dror, 2020).

1.10 Hierarchy of Expert Performance by Dror (2016)

Biases that arise from these various sources may affect decision-making processes, either by biasing the forensic observations or the conclusions. The Hierarchy of Expert Performance (HEP) provides an eight-level framework that distinguishes between the observation and conclusion elements in decision-making. The eight levels of HEP also address two other elements: a) reliability and biasability, and b) between and within experts (I. E. Dror & Murrie, 2018).

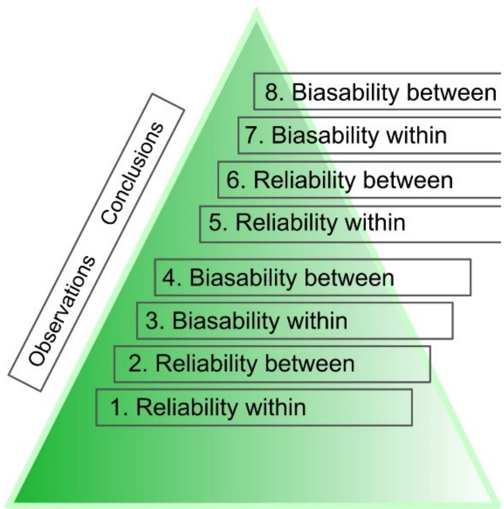

Figure 2 Dror's (2016) Hierarchy of Expert Performance (HEP) framework

Level 8 concerns the question of how different experts who are considering identical data are biased by irrelevant contextual information. In forensic science, examples of irrelevant contextual information include the suspect’s confession or whether other lines of evidence suggest that the suspect is the defendant. For example, a study showed that examiners exposed to irrelevant information that implicated a suspect could not dismiss the suspect from the suspect pool. In contrast, the examiners who were not exposed to the information did not reach the same conclusion (Dror & Hampikian, 2011).

Level 7 concerns whether the same expert reaches the same conclusion when the same case is presented with an irrelevant biasing context. A study showed that experts did not always reach the same conclusion when the same fingerprints were presented with different contextual information (Dror & Rosenthal, 2008).

Level 6 concerns whether experts examining the same information would reach the same conclusion, even without any contextual information that potentially biases the conclusion. Studies showed that when examining forensic evidence, such as fingerprints and DNA, experts will reach a spectrum of conflicting conclusions even if they examine the same evidence, free of any potentially biasing information (Dror & Hampikian, 2011; Dror & Rosenthal, 2008).

Level 5 focuses on whether the same expert would reach the same conclusion when repeatedly considering the same data. A study showed that the same fingerprint experts will not reach the same conclusion about 10% the time when repeatedly examining the same fingerprints (Ulery et al., 2012).

Level 1 through 4 examines the same issue concerning: a) reliability and bias ability, and b) between and within experts, in the level of observations. For example, studies showed that fingerprint examiners pay greater attention to different data depending on irrelevant contextual information (Dror et al., 2011; Earwaker et al., 2015). Furthermore, experts observe different data within the same evidence, between and within themselves, even when there is no irrelevant contextual information.

1.11 Forensic Psychiatry and Cognitive Biases in Insanity Evaluations

Forensic psychiatry is a branch of forensic science that is relevant to the discussion. Forensic psychiatrists have a role in the judicial process by examining the defendant's mental status to assess insanity. The psychiatrists are tasked with evaluating the mental status and determining whether the defendant was in a suitable mental status when the crime was committed does the defendant fall under the category of insanity (Melton et al., 2007). Psychiatrists are also tasked to evaluate a defendant whose competence is doubted to handle personal affairs, such as the ability to execute legal documents, or they may be asked to testify for the prosecution in a criminal case or the plaintiff in a civil suit or asked to determine the competency of the suspect and their ability to stand (Saferstein, 2018). The result of psychological testing usually supports the assessment of the defendant's mental competencies; sometimes, neurologic examination for organic brain dysfunction may also be used to support the psychiatric examination (Eckert, 1996). The legal status of criminal insanity, which is a determining factor to the competency of a defendant to stand trial, is based on intent and the responsibility of the act lies in the presence of free will of a person; the core of the status lies in 2 ideas where the defendant must be determined to suffer from a defect of the mind and as a result of the defect, the defendant does not have the knowledge on the nature and quality of the act that they did (Ogloff et al., 1993). Hence, to determine insanity, forensic psychiatrists are asked to provide an explanation that proves a person's insanity based on four questions: 1. Was the defendant suffering from a defect of reason resulting from a disease of the mind? 2. Did the defendant know the nature of the act? 3. Did the defendant know the quality of the act? 4. Did the defendant know that he was doing wrong at the time of committing the unlawful act? (Eckert, 1996). An example case illustrating the role of a forensic psychiatrist was done in a study concerning cases of maternal filicide — child murder done by mothers. Hospital records of 39 severely mentally ill mothers adjudicated Not Guilty by Reason of Insanity (NGRI) for filicide were analyzed to identify this event's precursor characteristics and suggest prevention strategies. During the analysis, it was found that 72% of the mothers had a history of mental health treatment. 69% of the mothers were experiencing auditory hallucinations, more often the commanding type, and 49% suffered depression at the time of the offense. 38% of the filicides occurred during pregnancy or postpartum, and many had a history of postpartum psychosis. 72% of the mothers had also experienced considerable developmental stressors, such as the death of a parental figure or incest. Motives behind the filicide were predominantly "altruistic" (the murder is done out of love) or "acutely psychotic" (occurred during psychosis, without rational motive) (Friedman et al., 2005).

Among all the forensic referrals, insanity referral is the most common type of referral where the decision-making criteria are ambiguous. For an insanity evaluation (also called "criminal responsibility" or "mental state at the time of offense" evaluations), the expert is tasked with reconstructing the thought processes and behaviors of the defendant before and during the occurrence of the alleged offense (Melton et al., 2007; Packer, 2009). Not only that the process of reconstruction examinations relies heavily upon inference by integrating the defendant's clinical history with collateral data, but there are also no set standards for how these evaluations should be conducted or how the report should be structured (Melton et al., 2007). The heavy reliance on inference and the lack of standard rules create a bigger opening for biases to influence judgments than a more structured referral question (Neal, 2018).

1.12 Reliability of Forensic Psychiatry Experts

Beyond reconstruction, different forensic opinions could be reached between experts due to the powerful influence of source variability and biases (Guarnera et al., 2017). Interrater reliability measures the level of agreement between numerous independent raters. Field reliability, or the inter-rater reliability of practitioners working in ordinary practice settings typical of real-world employment, is particularly important in forensic psychology (Wood et al., 2016). In general, the field reliability of forensic opinions is either unknown or far from perfect.

For instance, a recent meta-analysis found that disagreements between pairs of independent evaluators evaluating the same defendant occurred in 15% to 30% of cases involving adjudicative competency examinations, which is one of the most popular forensic psychology methods (Guarnera et al., 2017). The justification for these results is that forming a forensic opinion is a difficult task. For instance, legal sanity assessments demand that therapists judge a defendant's mental state at the time of the offense, which may have been months or years before, using sparse and frequently inconsistent information. Complex decision processes involving the integration of various sources of data tend to settle at fair to moderate dependability rates, according to a survey of a variety of medical and psychological procedures (Guarnera et al., 2017). This dependability rate contrasts the rather agreement coefficient of simple object counts (e.g., counting decayed or missing teeth) or physical measurements (e.g., measuring organ size on an ultrasound), where reliability tends to be higher, with coefficients higher than .90 (Meyer et al., 2010). Divergent forensic opinions appear to be influenced by patterns of consistent individual variations among evaluators rather than mere inaccuracy or random variance. For example, evaluators' base rates of finding defendants incompetent or insane appear to vary substantially, even when all evaluators in the sample rated defendants selected randomly but still from the same population. For example, in a sample of 59 clinicians conducting a total of 4,498 evaluations of legal sanity, seven clinicians found zero defendants insane, while three clinicians found 50% of all defendants insane (Murrie & Warren, 2005). Similarly, some evaluators assign consistently higher or lower PCL-R scores than others, even when there are no apparent differences among examinees that might explain these scoring trends (Boccaccini et al., 2008). Another contributor among forensic evaluators is the amount of variety and the lack of structured clinical assessment tools. While there are more tools and agreements on what is considered proper practice than even a decade ago, forensic psychologists still vary significantly in their performance throughout any given forensic examination (Heilbrun & Brooks, 2010). For example, a study discovered that 74% of forensic clinicians in a large worldwide sample employed at least one structured assessment instrument in their most recent evaluations. In contrast, 26% relied only on clinical judgment (Neal & Grisso, 2014). The 434 clinicians in the sample reported using 286 different instruments, many of which had questionable reliability or validity. Furthermore, even within a certain type of examination, the sources of information clinicians reported (e.g., medical records, judicial records, educational records, collateral interviews, psychological testing) varied greatly (Guarnera et al., 2017).

1.13 Forensic Psychiatrists and HEP

The HEP framework also applies to forensic psychology and psychiatry to measure the effect of bias in those fields. If we were to examine the biasability between experts' conclusions (level 8), there is one irrelevant contextual information that often plagues the conclusions made by the expert, which is the side that is retained by the experts, creating a vulnerability for adversarial allegiance (Murrie & Boccaccini, 2015; Neal, 2016). In a study where the experts were made to believe that they were working with either a prosecutor or a defense attorney, PCL-R and the static-99R measures showed that experts working for the prosecution assigned significantly higher scores. Those who thought they worked for the defense assigned lower scores (Murrie et al., 2013). Potentially biasing information goes beyond which sides retain the experts and other information, such as information about the defendant (case description, race, ethnicity, and others) (Dror & Murrie, 2018).

Next, reliability between experts' conclusions (level 6) is usually measured by examining whether experts reach the same conclusion in the same case. Studies have reported that experts' agreements have been poor when they use unstandardized procedures when assigning diagnoses (Aboraya et al., 2006). Regarding legal insanity, a study revealed that independent experts reached different conclusions 45% of the time (Gowensmith et al., 2013). Furthermore, during insanity evaluations with the goal of conditional release, experts reached different conclusions in 47% of the cases (Neil Gowensmith et al., 2017).

Within-level biasability (level 7) and reliability (level 5) pose a rather tricky problem. Unlike forensic science experts, forensic psychologists would be more likely to recognize that they are examining the same defendants. While interrater reliability is unlikely to be assessed due to the difficulty of ensuring that the clinician reviewed the same defendant "blindly". Nevertheless, these types of studies are critical to minimize potential biases in exploring the reliability of forensic conclusions (Dror & Murrie, 2018).

Observational levels (levels 1-4) also pose their difficulties. The distinction between observation and conclusion could be quite muddled in the forensic psychology and psychiatry. Sometimes, it seems clear what is an observation and a conclusion; however, other times, the distinction is not that clear-cut (Dror & Murrie, 2018).

1.14 Insanity Evaluation in Court and How to Improve it

In the penal sphere, holding a person criminally liable requires proof that they performed the act and that there was the intent behind it. Proving the presence of the intent during the time the crime was committed is especially important; this is due to the diminished capacity doctrine – the question of whether the defendant's mental state negates an element of the crime (Phillips & Woodman, 2007). In some circumstances, the defendant may be declared mentally ill at the time of the crime due to organic, neurologic, or psychiatric disturbances, hence the presence of accountability becomes partial or nonexistent (Scarpazza et al., 2018; Scarpazza, Zampieri, et al., 2021). The central role of forensic psychological testimonies for NGRI cases has been receiving increased criticisms in recent years; among those are concerning the poor inter-rater reliability of diagnostic categories, presence of intentional or unintentional biased reporting, and the lack of standardized methods and guidelines for procedures (Beckham et al., 1989; Guarnera et al., 2017). Forensic psychological evaluations are considered important components in cases involving mental health issues and are subject to strict judicial assumptions of evidentiary reliability and objectivity. Therefore, when reviewing the eligibility of the testimony, judges usually adhere to the Daubert standard (Goodman-Delahunty, 1997; Gowensmith & McCallum, 2019). Following the Daubert criteria, the validity of evidence must consider whether the theory, principle, or technique is testable (e.g., has a known possible error rate), has been subjected to peer review (e.g., in a scholarly peer-reviewed journal), and has general acceptance in the relevant scientific community (Daubert v. Merrell Dow Pharmaceuticals). The criteria push the discipline of psychology to now endorse strongly that information and services must be evidence-based (Woody, 2016).

Possible ways to improve the reliability of insanity evaluations are either to develop a more structured methodology or a way to measure concrete evidence that could indicate a causal link to insanity. In recognition that biological processes are at some level implicated in the development of criminal behavior, it would make sense to consider the idea that biological factors are relevant to understanding crime (Rafter, 2008; Raine, 2002). Some studies do show that alterations to the brain, such as the frontal lobe (Brower & Price, 2001; Williams et al., 2018), could increase the risk of criminal behaviors (Pietrini et al., 2000; Romero-Martínez et al., 2019; Sajous-Turner et al., 2020; Scarpazza, Zampieri, et al., 2021; Schug et al., 2010). As neuroscientific knowledge is getting extremely relevant for assisting in psychiatric assessments of criminal responsibility, one of the tools that have the potential to resolve reliability issues of traditional insanity evaluation is structural neuroimaging (sNI). The concept is that sNI would give additional biological data that, when combined with traditional neurological, psychiatric, and neuropsychological tests, would enlighten the court about the defendant's liability, moving a step further toward a verdict beyond any reasonable doubt (Scarpazza, Miolla, et al., 2021). According to the literature, sNI evidence should not be utilized alone and cannot be used to explain the reason for a violent crime solely. Instead, results obtained through sNI should be seen as a correlating factor for mental illness, in which the symptoms are causally connected to the crime. As a result, the criteria for responsibility are now behavioral and should stay such (Scarpazza, Zampieri, et al., 2021). With this important concept in mind, structural neuroimaging has been successfully applied in some insanity cases. For example, a 64-year-old male was arrested on a charge of pedophilia. He had no previous relevant medical, neurological, or psychiatric history; however, he showed symptoms of optic chiasm compression and frontal lobe dysfunction. Upon an MRI scan examination, it was found there was a 4×3cm Clivus Chordoma that was pressing on the pituitary gland and compressed the OFC, the optic chiasm, and the hypothalamus. Upon resectioning of the of the tumor, all pedophilic urges and other behavioral, neurological, and neuropsychological abnormalities dissipated (Sartori et al., 2016). This case highlighted the possibility that brain alteration could cause acquired pedophilia.

Another example was the case of a 55-year-old male nurse who was charged with multiple counts of sexual abuse. When the insanity assessment was conducted, experts from the judge and the defense agreed that the defendant showed the presence of narcissistic personality features that were clinically evident. However, experts on both parties could not agree on whether the traits were severe enough to be classified as psychopathological. When a structural MRI scan was performed, the defendant showed a bilateral increase in putamen gray matter volume This volume increase in the putamen is relevant because this increase had been observed across many psychiatric disorders; therefore, high specificity for psychopathology was found in the defendant's brain (Scarpazza, Zampieri, et al., 2021). It is crucial to remember that the structural neuroimaging scans here were not used as evidence of a direct cause of the charged crime or as a diagnostic material. However, it was used in conjunction as indirect support for the presence of psychopathology.

1.15 Aims of the Study

Although using sNI in court appears to be a hopeful step forward in lowering the uncertainty of forensic specialists' opinions, it is not without hazards and biases. One such bias is the assumption that neuroimaging results are objective, which is not necessarily the case (Brady, 2017). For example, when examining radiology errors, it was found that 20-40% of the errors are due to cognitive errors (i.e., failure to report the significance of abnormalities found) and 60-80% of the errors are due to perceptual error (i.e., failure to identify the abnormalities) (Brady, 2017; Brady et al., 1994; Lee et al., 2013). Moreover, some concerns were raised that the usage of neuroscience and neuroimaging have blurred the distinction between legally responsible and not legally responsible (Gurley & Marcus, 2008). For example, does the legal culpability of the defendant change if his/her brain is different from the norm? A study showed that mock jurors are more likely to render verdicts of NGRI when presented with neuroimaging data as it gives tangible proof of a disorder, highlighting the power of neuroimaging as a deciding factor on insanity (Gurley & Marcus, 2008). However, it is once again crucial to remember that the stance we take is that the usage of sNI is used as the supporting material of a diagnostic and not as the diagnostic itself.

Unlike those of medical disorders, diagnoses of psychiatric disorders, and the reliability of the accompanying sNI data, depend on experts who oversee the analysis of data and formulation of their interpretation. As a result, following previous research on the death certificate information discussed above, we sought to see if the availability of information could alter judgments made by neuroradiologists. We wanted to see whether the presence of irrelevant contextual information would alter the experts' judgment when examining the sNI data in deciding whether the imaging results could contribute to the defendants' insanity. To this aim, we provided neuroradiologists with the description of four different cases. For each case, we presented both the MRI and a description. Critically, while the MRI was the same across three different surveys we created, the information provided regarding the "patient" differed: sometimes only the demographic data were reported (i.e., age and gender), and sometimes demographic data were coupled with clinical data (i.e., age and gender + disease), and sometimes also forensic information was provided (i.e., age and gender + disease + crime committed). Neuroradiologists were asked to identify the brain abnormalities in each patient and to express their opinion on whether these abnormalities could be considered clinically relevant or not (i.e., could be the cause of the symptoms or not). We hypothesized that including clinically irrelevant information (i.e., the crime committed) would create a bias in neuroradiologists and would decrease the percentage of them that consider brain abnormalities that are potentially clinically relevant.

2. Methods and Procedure

2.1 Method

The experiment has a between-subjects structure. Participants were asked to respond to one out of four surveys. Each survey included the description of the same four cases, providing each different case amount of information.

2.2 Participants

All participants in this study were neuroradiologists that came from Associazione Italiana Neuroradiologia / Italian Society of Medical and Interventional Radiology (SIRM). The SIRM is one of the major non-profit Italian and European scientific societies founded in 1913, located in Via della Signora, 20122 Milano. A total of 33 neuroradiologists participated, which consisted of 21 females (63.6%) and 12 males (36.4%). Among those participants, 5 were between the age of 30-35 (15.2%), 6 were between the age of 36-40 (18.2%), 5 were between the age of 41-45 (15.2%), 9 were between the age of 46-50 (27.2%), 3 were between the age of 51-55 (9.1%), 2 were between the age of 56-60 (6%), and 3 were between the age of 61-65 (9.1%). The participants were collected through an email containing the link to the survey.

2.3 Materials

Surveys are created and spread using Qualtrics (Qualtrics, 2020) , a software created as a web-based survey tool used to conduct survey research, evaluation, and other data collection activities. In general, the survey presented four different cases, accompanied by short clinical descriptions and their respective MRIs. Although the same 4 cases were presented in each survey, the four surveys differ in the clinical information provided for each participant. For instance, the MRI of case 1 was accompanied by demographic information only in survey 1, demographic information and clinical description in survey 2, and demographic, clinical, and forensic information in survey 3. On the other hand, MRI of the cases, presented in a video format containing all the original sequences and plane acquisitions, was the same for each survey. All the surveys are done in Italian (translations are provided here to provide ease of reading). For each of the four cases, three different descriptions were created for the case: the first one including demographic data only, the second one including clinical data, and the third one also forensic data.

The presentation of the data for each survey follows this pattern:

- Survey 1: Case 1 – Demographic data; Case 4 – Demographic data; Case 2 – Demographic and medical data; Case 3 – Demographic, medical, and forensic data.

- Survey 2: Case 3 – Demographic data; Case 1 – Demographic and medical data; Case 4 – Demographic and medical data; Case 2 – Demographic, medical, and forensic data.
- Survey 3. Case 2 – Demographic data; Case 3 – Demographic and medical data; Case 1 – Demographic, medical, and forensic data; Case 4 – Demographic, medical, and forensic data.

2.4 Case Descriptions

Each of the presented cases is taken from real forensic cases. The three descriptions are of the following:

2.4.1 Case 1

- Demographic data: man, 33 years
- Demographic and medical data: man, 33 years, affected by a grave personality disorder and chronic substance abuse.
- Demographic, medical, and forensic data: man, 33 years, affected by a grave personality disorder, and chronic substance abuse. He has committed a double homicide, killing 2 of his friends

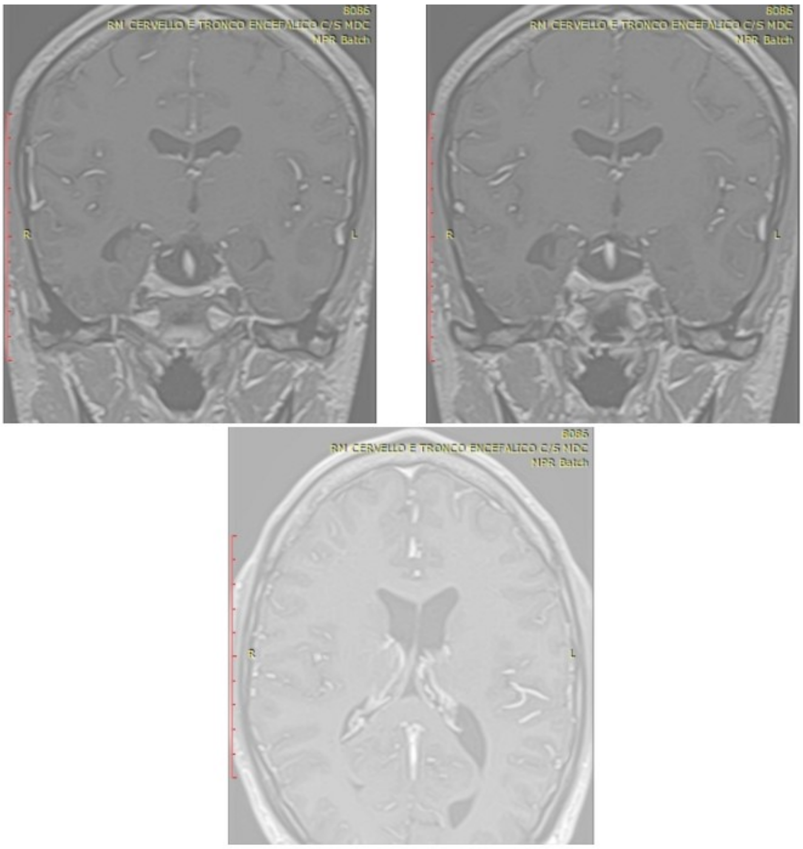

Figure 3 Brain abnormalities that are present in 55-years-old man. Top pictures show hippocampal atrophy that can be seen by the asymmetry of the hippocampi, especially where the left hippocampus appears normal, while the right appears atrophic. Bottom picture shows asymmetry of the ventricles, which can be seen by the left ventricle that is wider than the right. We point out that not only the temporal horn is more enlarged, but the whole left ventricle.

2.4.2 Case 2

- Demographic data: man, 54 years
- Demographic and medical data: man, 54 years, affected by psychotic-like symptoms and suspected mental retardation,
- Demographic, medical, and forensic data: man, 54 years, affected by psychotic-like symptoms and suspected mental retardation. He has committed murder, killing his wife and mother-in-law

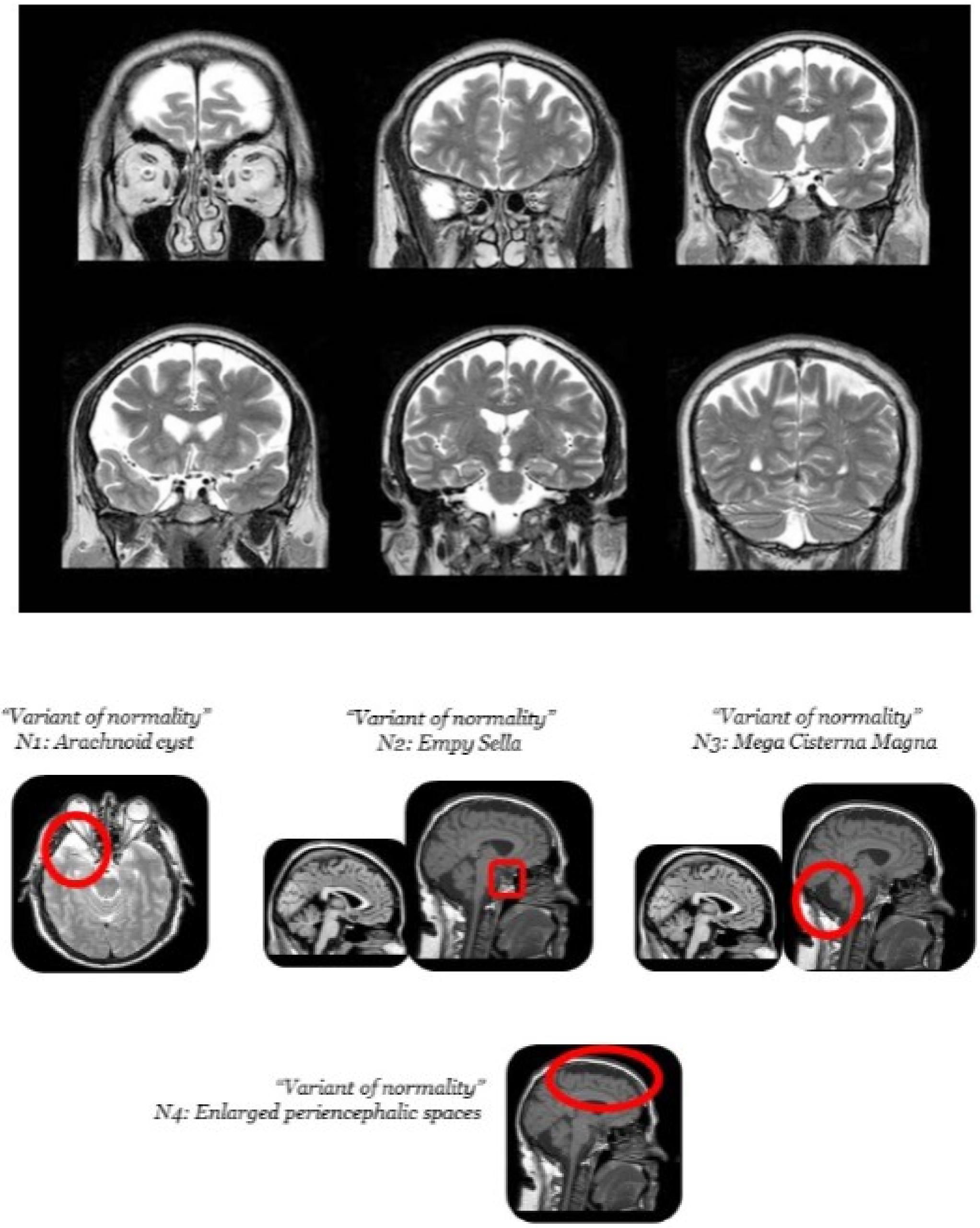

Figure 4 The four abnormalities of the 54-years-old man. i) an arachnoid cyst in the right temporal pole; ii) a mega cisterna magna (in the posterior part of the brain and in particular, in the cerebellum); iii) an empty sella; iv) enlarged periencephalic spaces around the frontal lobe.

2.4.3 Case 3

- Demographic data: man, 45 years
- Demographic and medical data: man, 45 years, has 15 years history of cocaine abuse.
- Demographic, medical, and forensic data: man, 45 years, has 15 years history of cocaineabuse. He has committed a homicide, killing his friend.

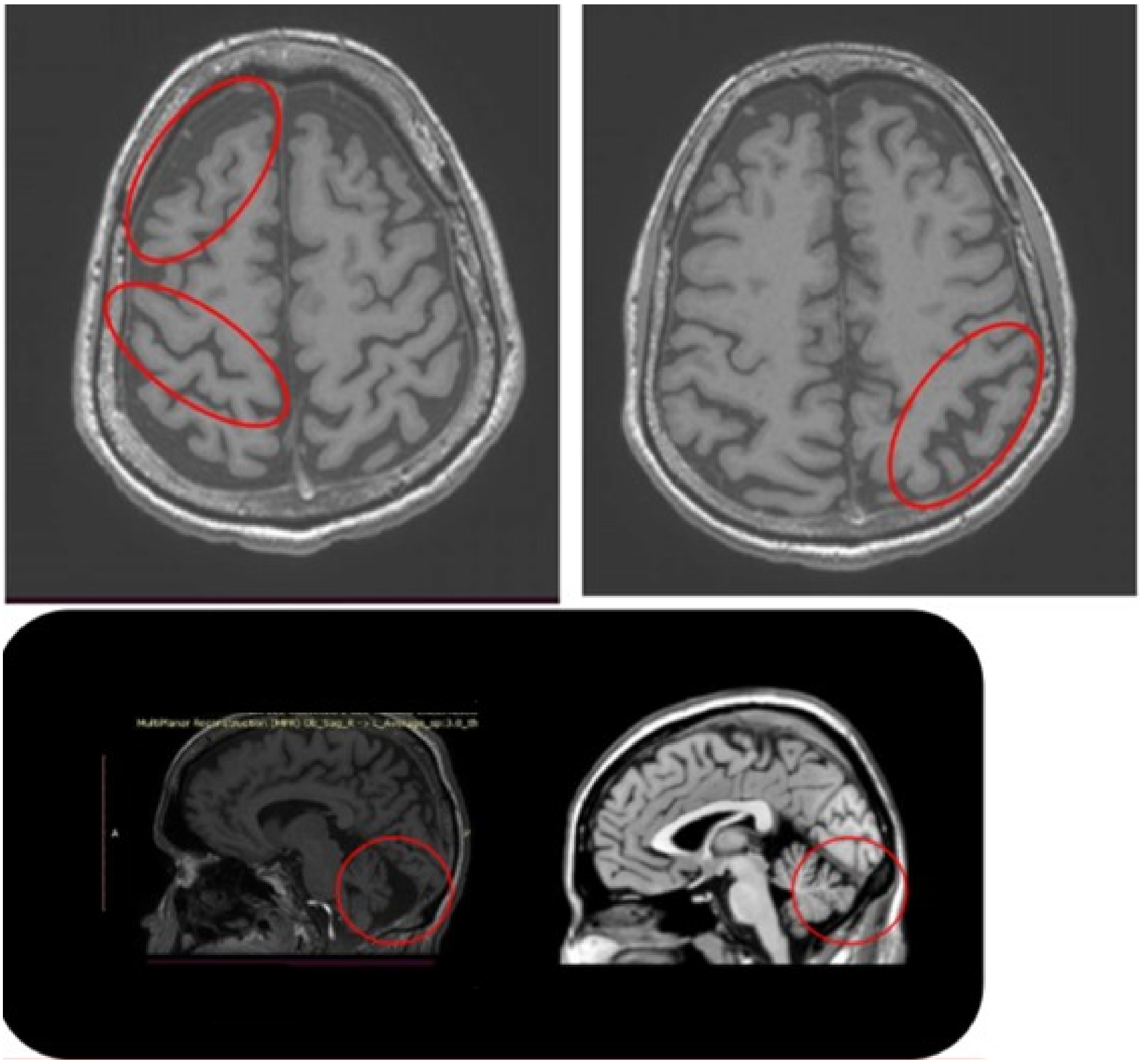

Figure 5 Abnormalities of the 45-years-old man. He has cortical atrophies localized in the frontal and parietal lobes

2.4.4 Case 4: the control condition for the study

- Demographic data: man, 60 years
- Demographic and medical data: man, 60 years, chronic alcoholic
- Demographic, medical, and forensic data: man, 45 years, chronic alcoholic. He has committed a homicide, killing his friend

2.5 Survey Structure

Four cases were presented for the survey. For each case, participants are first given the respective MRI scans and their accompanying information, depending on which survey they are assigned. Participants are then given three questions after they viewed the information given:

1. Did they detect anatomical anomalies in the case just seen? For this question, a binary choice of yes or no was given.
2. Could they briefly describe the type of anomaly (s) you may have encountered and the location they think you should report in your clinical report? A brief open-ended answer was required for this question.
3. Could the neuroanatomical anomalies found eventually have clinical relevance? Three choices are given for this answer: no, unsure, or yes.

These questions are repeated for all four cases. Afterward, participants are asked to provide relevant demographic information about their age, gender, profession duration, and location. During the process, participants are not allowed to return to the previous page of the survey once they have advanced to the next one.

2.6 Statistical Analyses

Statistical analyses were carried out using SPSS (IBM, 2022). Data were analyzed with non-parametric testing due to the small population sample size. Non-parametric tests used were the Kruskal-Wallis test (Kruskal & Wallis, 1952) and the Chi-square test (Tallarida & Murray, 1987).

3. Results

3.1 Gender

Across all three surveys, 12 out of 33 participants were males, and 21 out of 33 were females.

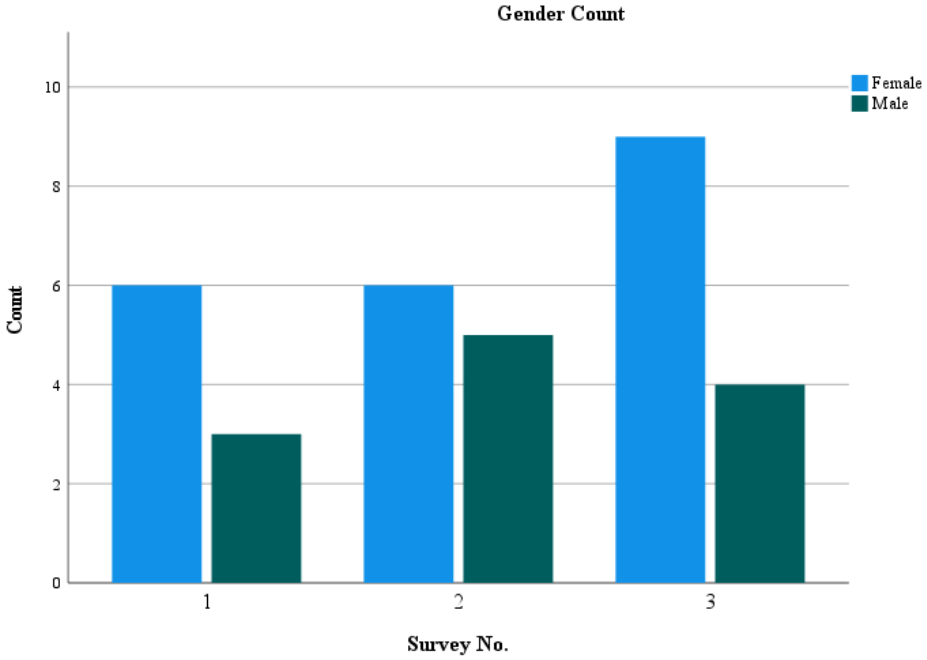

Figure 6 Chart showing the number of genders in all three survey conditions. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

Survey types and gender are independent of each other, in other words, the distribution of gender for each survey types are due to chance (X^2 (2, N = 33) = 0.604, p = 0.739).

3.2 Age and Work experience

The analyses from all three information conditions showed that the median of the age was 46-50 years old. With the youngest participants who were 30-35 years old and the oldest participants who were 61-65 years old.

The mean years of work experience for all three information conditions were 15.97 years (Md = 15, SD = 10.85). Furthermore, participants in the demographic information condition had a mean of 17.44 years of work experience. The participants in the demographic and medical information condition had a mean of 15.91 years of work experience. Lastly, the participants in the demographic, medical, and forensic information condition had a mean of 15 years of work experience.

A Kruskal-Wallis test showed that there was no statistically significant difference in age ($H(2) = 0.574$, $p = 0.750$) or work experience ($H(2) = 0.357$, $p = 0.837$) of participants who filled the three survey types.

3.3 Case 1

Case 1 describes a 33-year-old male with a history of cocaine abuse. Abnormalities present are hippocampal atrophy and ventricle asymmetry.

Firstly, the responses to whether the participants saw any anomalies when given the MRI scan and information were analyzed. There is an increase in the participants responding that they did not see any anomalies as more information was given to the participants. The proportion increased from 11.1% in survey 1, which provided demographic data only, to 53.8% in survey 3, which provided demographic, medical, and forensic data.

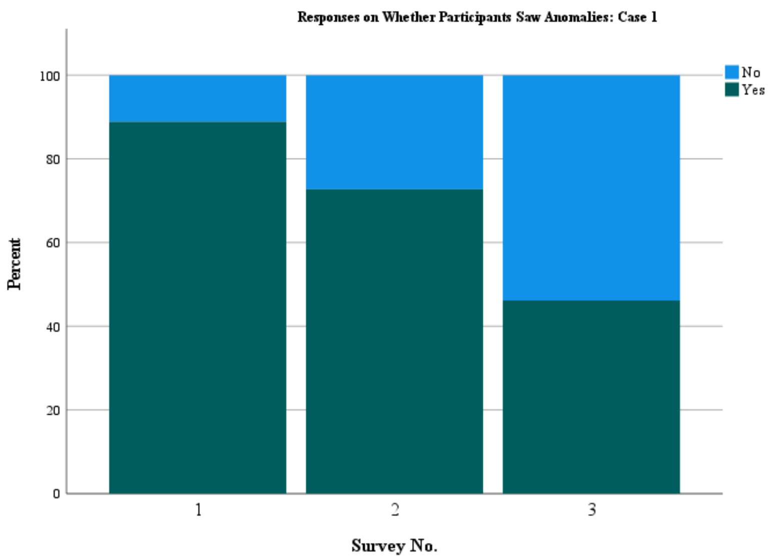

Figure 7.1 Number of Yes and No responses for all the information types of Case 1. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, there was no significant association between the different information being presented and whether the participants saw anomalies or not (X^2 (2, N = 33) = 4.643, p = 0.098). A likely explanation for the insignificance would be due to the small sample size.

Then the responses to whether the observed neuroanatomical anomalies could have any clinical relevance were analyzed. There is also an increasing pattern of participants responding that the anomalies did not have clinical relevance as the survey versions changed, especially from 11,1% in the demographic data condition to 76.9% in the demographic, medical, and forensic data condition.

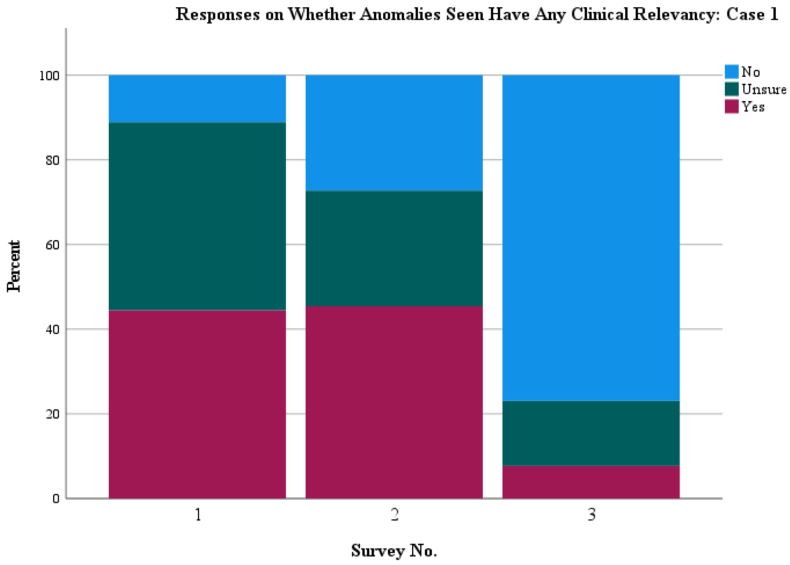

Figure 7.2 Responses regarding clinical relevance for Case 1. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

There was a significant association between the different information provided and whether the neuroanatomical anomalies perceived had any clinical relevance ($X^2(4, N = 33) = 11.590$, $p = 0.021$).

Finally, the responses on the number of anomalies listed by the participants were analyzed. For all three different types of information, participants listed one anomaly on average.

In addition, a Kruskal-Wallis test showed that the number of anomalies was not affected by the information given ($H(2) = 1.473$, $p = 0.479$). The mean ranks of the number of anomalies listed are 19.72 for the demographic information condition, 17.23 for the demographic and medical information condition, and 14.92 for the demographic, medical, and forensic information condition.

Particularly for this case, one conclusion can be reached. When forensic information is provided, participants are more likely to not see the abnormalities. However, when participants are able see the abnormalities, they are more likely to consider it clinically irrelevant.

3.2 Case 2

Case 2 describes a 54-year-old man, affected by psychotic-like symptoms and suspected mental retardation. He has four brain abnormalities, which are an arachnoid cyst in the right temporal pole, a mega cisterna magna, an empty sella, and an enlarged periencephalic spaces around the frontal lobe.

Analyses were done to determine whether the participants saw any anomalies after they are given the MRI scan and information. For all three versions, most participants agreed that they had observed some anomalies.

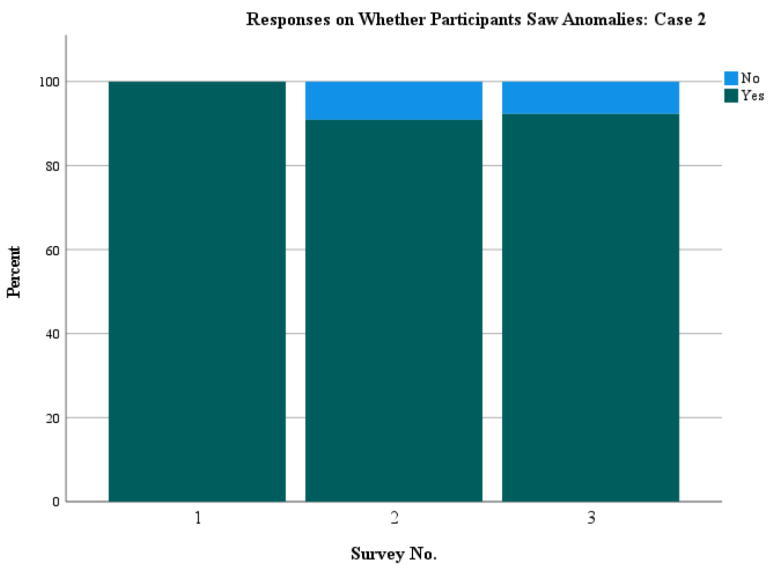

Figure 8.1 Number of Yes and No responses for all the information types of Case 2. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, there was no significant association between the different information provided and whether the participants saw anomalies or not (X^2 (2, N = 33) = 0.819, p = 0.664). A ceiling effect was observed here, as most participants can see the anomalies, causing the agreement response to be at a maximum. Hence, the different information types no longer influence the response.

Then, the responses to whether the neuroanatomical anomalies observed could have clinical relevance were analyzed. There is a decrease in agreements of participants responding regarding whether the anomalies they saw had a clinical relevance as different types of information are given, especially from 88.9% in the demographic data condition to 36.4% in the demographic and medical data condition to 38.5% in the demographic, medical, and forensic data condition.

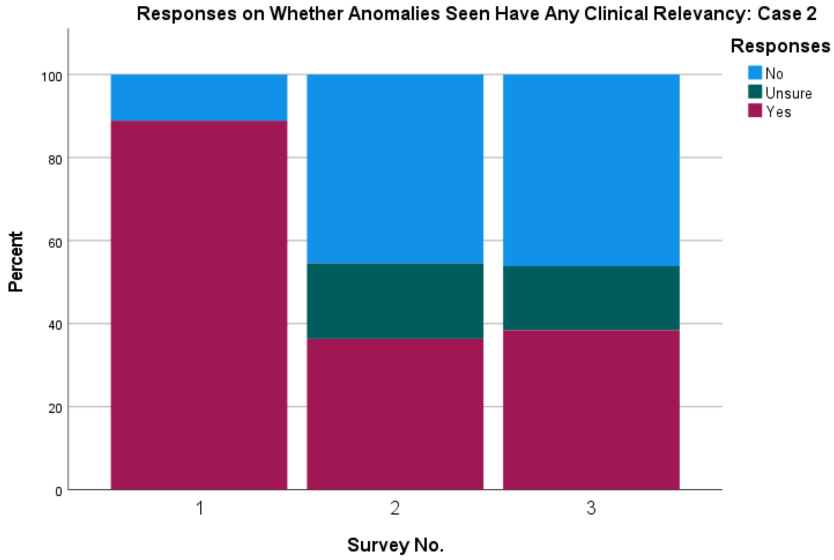

Figure 8.2 Responses regarding clinical relevance for Case 2. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, there was no significant association between the different information given and whether the neuroanatomical anomalies perceived had any clinical relevance ($X^2(4, N = 33) = 7.070, p = 0.132$). A likely explanation for this could also be due to the small sample size.

Finally, the analyses of the number of anomalies the participants listed showed that for all three types of information condition, participants listed two anomalies on average.

In addition, a Kruskal-Wallis test showed that the number of anomalies was not affected by the type of information given ($H(2) = 0.602, p = 0.740$). The mean ranks of the number of anomalies listed are 15 for the demographic information condition, 17.32 for the demographic and medical information condition, and 18.21 for the demographic, medical, and forensic information condition.

In conclusion, regarding the demographic, medical, and forensic data condition in case 2, participants can see the anomalies that were shown. However, they were more reluctant to consider these anomalies to be potentially relevant clinically.

3.3 Case 3

Case 3 describes a 45-year-old man, has 15 years history of cocaine abuse. He has cortical atrophies localized in the frontal and parietal lobes

Responses to whether the participants saw any anomalies when given the MRI scan and information were analyzed. Comparison between survey types showed that the participants reported an increase in seeing anomalies, from 44.4% when the participants were given demographic data only to 81.8% when also given medical information and 61.5% when also given medical and forensic information.

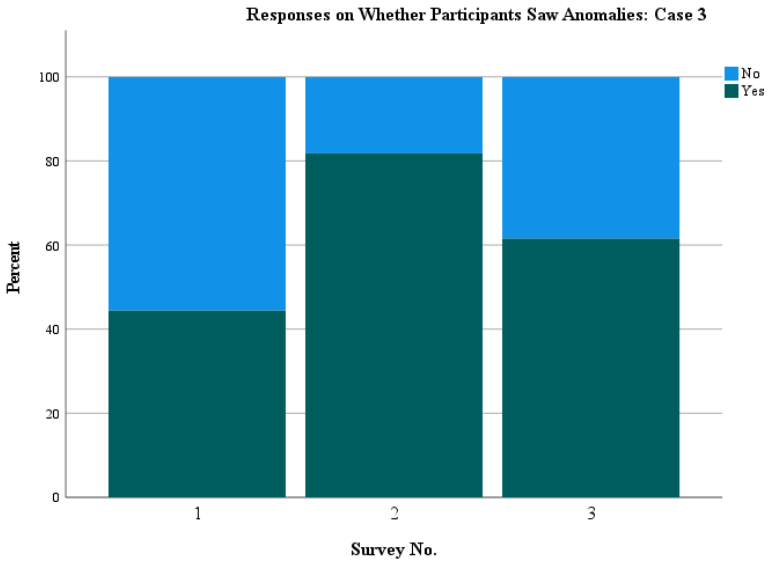

Figure 9.1 Number of Yes and No responses for all the information types of Case 3. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, there was no significant association between the survey types and whether the participants saw anomalies or not ($X^2(2, N = 33) = 3.029, p = 0.220$).

Then, from the responses to whether the neuroanatomical anomalies the participants observed could have any clinical relevance, there was an increase in agreements of participants' responses regarding whether the anomalies they saw had a clinical relevance when compared between survey types, especially from 11.1% when given demographic information only to 54.5% when also given medical information and 46.2% when also given medical and forensic information.

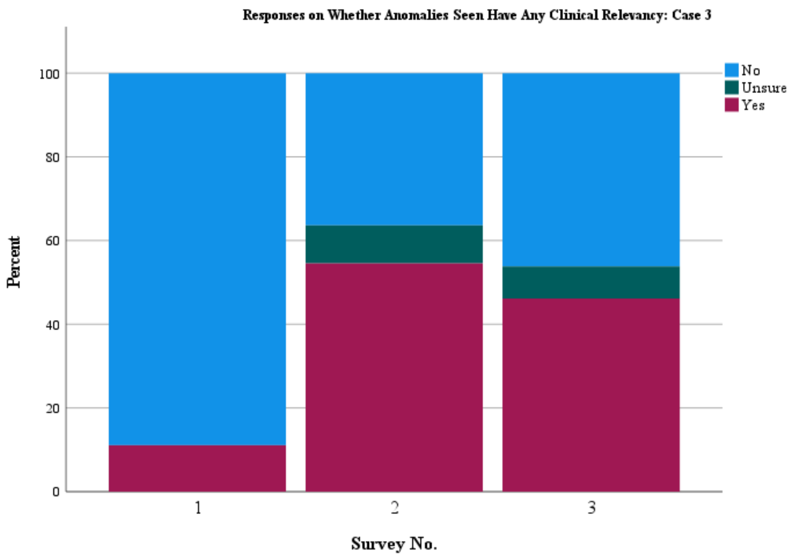

Figure 9.2 Responses regarding clinical relevance for Case 3. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, there was no significant association between the information given and whether the neuroanatomical anomalies perceived had any clinical relevance ($X^2(4, N = 33) = 6.169$, $p = 0.187$)

Finally, subjects listed one anomaly on average for all three-information condition.

A Kruskal-Wallis test showed that the number of anomalies was not affected by the type of information given ($H(2) = 2.634$, $p = 0.268$). The mean ranks of the number of anomalies listed are 13.78 for the demographic information condition, 20.23 for the demographic and medical information condition, and 16.50 for the demographic, medical, and forensic information condition.

3.4 Case 4

Case 4 describes a 60-year-old man, chronic alcoholic. He has no brain abnormalities and serves as a control condition.

Through observation alone, there was a decrease in the rate of subjects' agreement to whether they did not see any anomalies across different information condition The decrease went from 77.8% in the demographic data condition to 46.2% in demographic, medical, and forensic data condition.

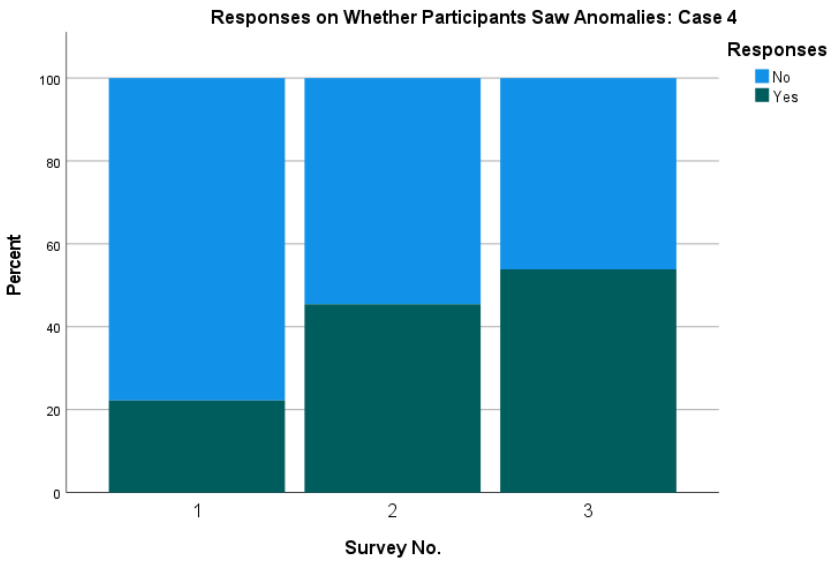

Figure 10.1 Number of Yes and No responses for all the information types of Case 4. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

However, further analysis showed there was no significant association between the survey types and whether the subjects saw anomalies or not (X^2 (2, N = 33) = 2.239, p = 0.326). The MRI scans for this case was reported to be affected by motion artifacts. Hence, the responses stating that abnormalities are present could be signal abnormalities instead of an actual neuroanatomical one.

Next, subjects reported, in all three survey versions, that they mostly did not see any anomalies deemed to have clinical relevance and this applied to all three different information conditions.

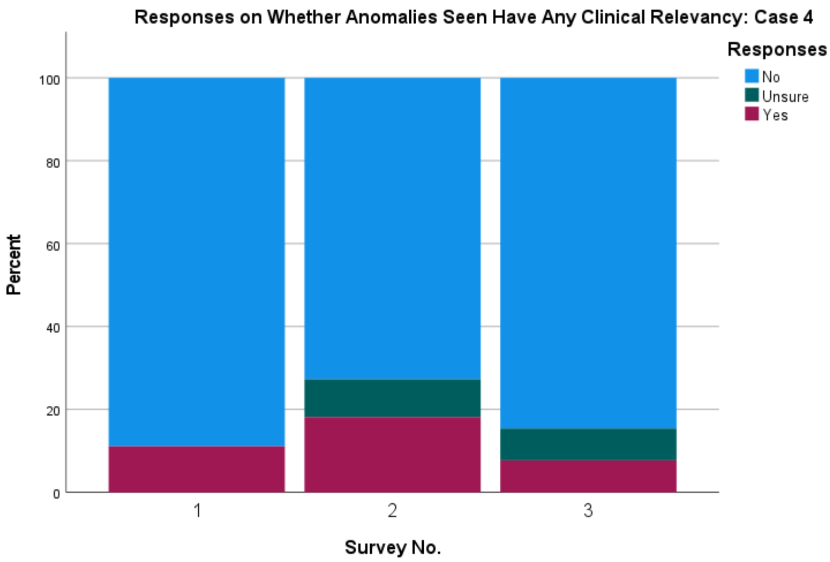

Figure 10.2 Responses regarding clinical relevance for Case 4. Survey 1 refers to the demographic information. Survey 2 refers to the demographic and medical information. Survey 3 refers to demographic, medical, and forensic information.

There was no significant association between the information conditions and whether the neuroanatomical anomalies perceived had any clinical relevance ($X^2(4, N = 33) = 1.499$, $p = 0.827$).

Finally, for all three surveys, subjects mainly did not report any number of anomalies.

A Kruskal-Wallis test showed that the number of anomalies was not affected by the type of information given ($H(2) = 0.836$, $p = 0.658$). The mean ranks of the number of anomalies listed are 14.78 for the demographic information condition, 17.86 for the demographic and medical information condition, and 17.81 for the demographic, medical, and forensic information condition.

4. Discussion

As stated above, following previous research on death certificate information, we sought to see if the availability of information, whether relevant or not, would alter the experts' judgment when examining the sNI data in deciding whether the imaging results could contribute to the defendants' insanity. We understand that there are limitations to the study, an evident one being the modest sample size causing results to be more often insignificant. Due to this, any definitive conclusions from the current study cannot be taken. Although, this does not deny the possibility that contextual information, in this case, the information type presented, does not affect the judgment of neuroradiologist experts when analyzing brain scans.

According to the HEP, two basic properties of an expert's decision making is biasability and reliability (I. E. Dror, 2016). The setting of this study allows us to examine the reliability and biasability between experts, both in the level of observation and conclusion. Biasability refers to the ability to make decisions based on relevant information without being biased by irrelevant contextual information. Reliability refers to the ability to make decisions consistently, even without irrelevant biasing information. As the presentation of sNI scans is the same for all the conditions, the only variable that differs is the presentation of information type; this allows us to examine the effect of information type on the biasability and reliability of experts' decision-making.

On the observational level, the biasability effect of contextual information, especially forensic information, on neuroradiologists' judgment is particularly visible in case 1. Despite sNI scans showing abnormalities (hippocampal atrophy and ventricle asymmetry), participants presented with the information on age and gender judged that there were anomalies present when reviewing the brain scans (88.9%). However, when they were also given the forensic information (the crime that the defendant was charged with), participants affirmative response rate strongly declined (46.2%).

The reliability of neuroradiology experts on an observational level can also be examined. In case 1, inconsistencies appeared when experts were given demographic, medical, and forensic information. Experts' responses to whether there is the presence of brain anomalies became divided, with slightly more than half (53.8%) responding "no" and the other (46.2%) responding "yes". Furthermore, in the demographic information condition for case 3, there was also a split in between the rate of agreements of responses between experts. More than half of the participants (55.6%) responded that they did not see the presence of anomalies, and the other half (44.4%) responded that there were anomalies present.

On the conclusion level, the biasability of expert judgment can be examined, particularly in case 1. When participants were only given the demographic information, only a minority of participants (11.1%) concluded that the observed anomalies did not have any clinical relevance to the case. However, when forensic information was provided, the conclusion that the anomalies observed did not have any clinical relevancy became the majority (76.9%).

Lastly, the reliability of neuroradiologist experts on the conclusion level can also be examined in case 1. In the demographic information condition, the consistency of experts' conclusions is put in question as the conclusion on whether the brain anomalies observed could have clinical relevancy are split between "yes" (44.4%) and "unsure" (44.4%).

Reliability is an important parameter to examine in decision-making as there needs to be some degree of consistency to examine a phenomenon. If decisions on the same evidence are different each time, then there is no pattern in the decisions (I. E. Dror et al., 2021). Neuroradiologist examinations do not come with inbuilt labels on what constitutes the anomalies, and the interpretation of results is not an easy task. A previous study indicated that approximately 4% of radiologic interpretations in daily practice contain errors (Borgstede et al., 2004; Brady, 2017; Lee et al., 2013). The unreliability in neuroradiology decision-making could be attributed to the idea that radiologists' conclusions rely on several premise (e.g., available clinical information, statistical likelihood, relevant history, and previous imaging). Discrepancies on the premises could contribute to an erroneous judgment (Brady, 2017). Furthermore, since the interpretation of the premises can potentially be different between individuals, it creates the possibility that a different radiologist might have come to a different conclusion based upon the same information (Dror et al., 2021). The reliance on the human analysis for interpreting the examination of radiology could explain the low reliability among neuroradiologist experts (Abujudeh et al., 2010).

Another issue comes when neuroradiological identification task increases in difficulty, for example, due to technical issues (e.g., movement artifacts, incorrect imaging protocols), workload fatigue (e.g., excess workload), or other cognitive factors (e.g., biases, inattention, or mental fatigues) (Brady, 2017; Lee et al., 2013). Due to cognitive processing, complex tasks allow more influence by top-down (conceptually driven) processing. Therefore, there is more leeway for contextual information to mediate how the data are processed and the conclusions are reached (Dror et al., 2021). Even if this study demonstrated that there is a possibility that contextual information can affect decision-making, this does not mean that all contextual information is irrelevant. The level of relevancy of contextual information varies from case to case and different decision types (Dror et al., 2021).

In this study, decisions on anomalies and their clinical relevancy should not be based on information such as the forensic description (Dror et al., 2021). This is because forensic information, such as the crimes the defendant committed, is non-medical and task-irrelevant in forensic decision-making. However, we can see that forensic description instead alters the decision-making when presented instead. This highlights the need to consider the potential impact of irrelevant contextual data on the formation of forensic conclusions.

Further research is needed to corroborate further to draw conclusive evidence regarding decision-making in experts and how information biases these decisions. However, it provides a good starting point for future progression of the study. It also provides a chance to consider any potential future directions and solutions to mitigate the issue of bias in decision-making. Strategies for minimizing radiological error could also be used in the context of forensic decision-making. For example, during the education of experts, knowledge of the bias and how it may affect decision-making could help experts become more aware of these biases (Brady, 2017; Lee et al., 2013). Another potential strategy is to rely on the collaboration of a radiographer for data collection and a radiologist for data interpretation instead of only relying on one radiologist. By including a radiographer in the data collection process, the radiologists' workload can be reduced, lessening their mental fatigue (Schneider et al., 2012). Also, suppose radiographers are given proper blinding (i.e., making diagnostic reports unavailable during data collection) (Brealey & Scally, 2001). In that case, this could increase the quality of data gathered by reducing any variables that may affect the observation and collection process. Data obtained by the radiographer are then passed to radiologists for the final interpretation. Lastly, the radiographer's final interpretation of sNI scans should be subjected to peer review, which is the systematic evaluation of performance using a structured process (Lee et al., 2013). Peer review aims to improve diagnostic accuracy and reduce human error in interpretation (Strickland, 2015).

Bibliography

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis. Psychiatry (Edgmont), 3(1), 41. /pmc/articles/PMC2990547/

Abujudeh, H. H., Boland, G. W., Kaewlai, R., Rabiner, P., Halpern, E. F., Gazelle, G. S., & Thrall, J. H. (2010). Abdominal and pelvic computed tomography (CT) interpretation: Discrepancy rates among experienced radiologists. *European Radiology*, *20*(8), 1952–1957. https://doi.org/10.1007/S00330-010-1763-1/TABLES/3

Ayyub, B. M., & Klir, G. J. (2006). Uncertainty modeling and analysis in engineering and the sciences. *Uncertainty Modeling and Analysis in Engineering and the Sciences*, 1–378. https://doi.org/10.1201/9781420011456/UNCERTAINTY-MODELING-ANALYSIS-ENGINEERING-SCIENCES-BILAL-AYYUB-GEORGE-KLIR

Beckham, J. C., Annis, L. v, & Gustafsont, D. J. (1989). Decision Making and Examiner Bias in Forensic Expert Recommendations for Not Guilty by Reason of Insanity. *Law and Human Behavior, 13*(1).

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, *20*(1), 115–129. https://doi.org/10.1002/ACP.1171

Blanchette, I., & Richards, A. (2009). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Https://Doi.Org/10.1080/02699930903132496*, *24*(4), 561–595. https://doi.org/10.1080/02699930903132496

Blanco, F. (2017). *Cognitive bias C Cognitive Bias*. https://doi.org/10.1007/978-3-319-47829-6_1244-1

Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). DO SOME EVALUATORS REPORT CONSISTENTLY HIGHER OR LOWER PCL-R SCORES THAN OTHERS?: Findings From a Statewide Sample of Sexually Violent Predator Evaluations. *Psychology, Public Policy, and Law*, *14*(4), 262–283. https://doi.org/10.1037/a0014523

Borgstede, J. P., Lewis, R. S., Bhargavan, M., & Sunshine, J. H. (2004). RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. *Journal of the American College of Radiology : JACR*, *1*(1), 59–65. https://doi.org/10.1016/S1546-1440(03)00002-4

Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable? *Insights into Imaging*, *8*(1), 171–182. https://doi.org/10.1007/S13244-016-0534-1

Brady, A. P., Stevenson, G. W., & Stevenson, I. (1994). Colorectal cancer overlooked at barium enema examination and colonoscopy: A continuing perceptual problem. *Radiology*, *192*(2), 373–378. https://doi.org/10.1148/RADIOLOGY.192.2.8029400

Brealey, S., & Scally, A. J. (2001). Bias in plain film reading performance studies. *British Journal of Radiology*, *74*(880), 307–316. https://doi.org/10.1259/BJR.74.880.740307

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673–681. https://doi.org/10.1037/0022-3514.42.4.673

Brower, M. C., & Price, B. H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: A critical review. In *Journal of Neurology Neurosurgery and Psychiatry* (Vol. 71, Issue 6, pp. 720–726). https://doi.org/10.1136/jnnp.71.6.720

Buckhout, R., Figueroa, D., & Hoff, E. (1975). Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Bulletin of the Psychonomic Society*, *6*(1), 71–74. https://doi.org/10.3758/BF03333151

Carbon, C. C. (2014). Understanding human perception by human-made illusions. *Frontiers in Human Neuroscience*, *8*(JULY), 566. https://doi.org/10.3389/FNHUM.2014.00566/BIBTEX

Daubert v. Merrell Dow Pharmaceuticals, Inc. :: 509 U.S. 579 (1993) :: Justia US Supreme Court Center. (n.d.). Retrieved August 22, 2022, from https://supreme.justia.com/cases/federal/us/509/579/

Davis, N. (2017). The selfish gene. *The Selfish Gene*, 1–80. https://doi.org/10.4324/9781912281251

Deffenbacher, K. A. (2008). Estimating the impact of estimator variables on eyewitness identification: A fruitful marriage of practical problem solving and psychological theorizing. *Applied Cognitive Psychology*, *22*(6), 815–826. https://doi.org/10.1002/ACP.1485

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, *8*(JUL). https://doi.org/10.3389/FPSYG.2017.01153/FULL

Dror, I. (2015). Cognitive neuroscience in forensic science: Understanding and utilizing the human element. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 370, Issue 1674). The Royal Society. https://doi.org/10.1098/rstb.2014.0255

Dror, I. (2020). Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. In *Analytical Chemistry* (Vol. 92, Issue 12, pp. 7998–8004). American Chemical Society. https://doi.org/10.1021/acs.analchem.0c03002

Dror, I. E. (2016). A Hierarchy of Expert Performance. *Journal of Applied Research in Memory and Cognition*, *5*(2), 121–127. https://doi.org/10.1016/J.JARMAC.2016.03.001

Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., & Rosenthal, R. (2011). Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a “target” comparison. *Forensic Science International*, *208*(1–3), 10–17. https://doi.org/10.1016/j.forsciint.2010.10.013

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

Dror, I. E., & Murrie, D. C. (2018). A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law*, *24*(1), 11–23. https://doi.org/10.1037/LAW0000140

Dror, I. E., Scherr, K. C., Mohammed, L. A., MacLean, C. L., & Cunningham, L. (2021). Biasability and reliability of expert forensic document examiners. *Forensic Science International*, *318*, 110610. https://doi.org/10.1016/j.forsciint.2020.110610

Dror, I., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. S. (2021). Cognitive bias in forensic pathology decisions. *Journal of Forensic Sciences*, *66*(5), 1751–1757. https://doi.org/10.1111/1556-4029.14697

Dror, I., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences*, *53*(4), 900–903. https://doi.org/10.1111/J.1556-4029.2008.00762.X

Dugdale, J. R., & Eklund, R. C. (2002). Do not pay any attention to the umpires: Thought suppression and task-relevant focusing strategies. In *Journal of Sport and Exercise Psychology* (Vol. 24, Issue 3, pp. 306–319). https://doi.org/10.1123/jsep.24.3.306

Earwaker, H., Morgan, R. M., Harris, A. J. L., & Hall, L. J. (2015). Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science and Justice*, *55*(4), 239–247. https://doi.org/10.1016/j.scijus.2015.01.007

Eckert, W. (1996). Introduction to Forensic Sciences, Second Edition. In *Introduction to Forensic Sciences, Second Edition*. CRC Press. https://doi.org/10.1201/9781439821909

Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., Ballantyne, K., Searston, R. A., Thompson, M. B., Tangen, J. M., Kemp, R. I., & Martire, K. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science & Justice*, *57*(2), 144–154. https://doi.org/10.1016/J.SCIJUS.2016.11.005

Edwards, W. (2016). Dynamic Decision Theory and Probabilistic Information Processings1: *Https://Doi.Org/10.1177/001872086200400201*, *4*(2), 59–74. https://doi.org/10.1177/001872086200400201

Ellis, G. (2018). Cognitive Biases in Visualizations. In *Cognitive Biases in Visualizations*. Springer International Publishing. https://doi.org/10.1007/978-3-319-95831-6

Friedman, S. H., Hrouda, D. R., Holden, C. E., Noffsinger, S. G., & Resnick, P. J. (2005). Child Murder Committed by Severely Mentally Ill Mothers: An Examination of Mothers Found Not Guilty By Reason of Insanity. *Journal of Forensic Sciences*, *50*(6), 1–6. https://doi.org/10.1520/JFS2005132

Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, *11*(2), 133–142. https://doi.org/10.1007/S10339-009-0345-0/TABLES/1

Ganis, G., & Kosslyn, S. M. (2007). Multiple mechanisms of top-down processing in vision. *Representation and Brain*, 21–45. https://doi.org/10.1007/978-4-431-73021-7_2/COVER

Gershman, S. J., & Beck, J. M. (2017). *Complex Probabilistic Inference: From Cognition to Neural Computation*.

Glöckner, A., & Witteman, C. (2009). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making.

Https://Doi.Org/10.1080/13546780903395748, *16*(1), 1–25. https://doi.org/10.1080/13546780903395748

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Goodman-Delahunty, J. (1997). Forensic psychological expertise in the wake of Daubert. *Law and Human Behavior*, *21*(2), 121–140. https://doi.org/10.1023/A:1024874228425

Gowensmith, W. N., & McCallum, K. E. (2019). Mirror, mirror on the wall, who's the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations: *Https://Doi.Org/10.1177/0081246319835117*, *49*(2), 165–176. https://doi.org/10.1177/0081246319835117

Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2013). How reliable are forensic evaluations of legal sanity? *Law and Human Behavior*, *37*(2), 98–106. https://doi.org/10.1037/LHB0000001

Granhag, P. A., Strömwall, L. A., & Hartwig, M. (2005). Eyewitness testimony: tracing the beliefs of Swedish legal professionals. *Behavioral Sciences & the Law*, *23*(5), 709–727. https://doi.org/10.1002/BSL.670

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science*, *3*(2), 143–152. https://doi.org/10.1037/tps0000114

Gurley, J. R., & Marcus, D. K. (2008). The effects of neuroimaging and brain injury on insanity defenses. *Behavioral Sciences and the Law*, *26*(1), 85–97. https://doi.org/10.1002/BSL.797

Hammond, K. (2000). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. https://books.google.com/books?hl=en&lr=&id=VxRnDAAAQBAJ&oi=fnd&pg=PR9&dq=Human+Judgment+and+Social+Policy:+Irreducible+Uncertainty,+Inevitable+Error&ots=xwOnCLFraE&sig=A15_ouPLm9JOVf--N-W0Gn5JaZE

Heilbrun, K., & Brooks, S. (2010). Forensic Psychology and Forensic Science: A Proposed Agenda for the Next Decade. *Psychology, Public Policy, and Law*, *16*(3), 219–253. https://doi.org/10.1037/A0019138

Hoc, J. M., & Amalberti, R. (2016). Cognitive Control Dynamics for Reaching a Satisficing Performance in Complex Dynamic Situations: *Http://Dx.Doi.Org/10.1177/155534340700100102*, *1*(1), 22–55. https://doi.org/10.1177/155534340700100102

Islam, R., Weir, C., & del Fiol, G. (2014). Heuristics in managing complex clinical decision tasks in experts' decision making. *Proceedings - 2014 IEEE International Conference on Healthcare Informatics, ICHI 2014*, 186–193. https://doi.org/10.1109/ICHI.2014.32

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming Famous Overnight: Limits on the Ability to Avoid Unconscious Influences of the Past. *Journal of Personality and Social Psychology*, *56*(3), 326–338. https://doi.org/10.1037/0022-3514.56.3.326

Julian, R. D., Kelty, S. F., Roux, C., Woodman, P., Robertson, J., Davey, A., Hayes, R., Margot, P., Ross, A., Sibly, H., & White, R. (2011). What is the value of forensic science? An overview of the effectiveness of forensic science in the Australian criminal justice system project. *Australian Journal of Forensic Sciences*, *43*(4), 217–229. https://doi.org/10.1080/00450618.2011.610820

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

Knauff, M., & Wolf, A. G. (2010). Complex cognition: the science of human reasoning, problem-solving, and decision-making. *Cognitive Processing 2010 11:2*, *11*(2), 99–102. https://doi.org/10.1007/S10339-010-0362-Z

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583. https://doi.org/10.2307/2280779

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners. *Journal of Applied Research in Memory and Cognition*, *6*(4), 452–459. https://doi.org/10.1016/J.JARMAC.2017.09.001

Lee, C. S., Nagy, P. G., Weaver, S. J., Newman-Toker, D. E., Cs, L., Pg, N., Sj, W., & De, N.-T. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. *American Journal of Roentgenology*, *201*(3), 611–617. https://doi.org/10.2214/AJR.12.10375

MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, *8*, 189–217. https://doi.org/10.1146/ANNUREV-CLINPSY-032511-143052

Magnussen, S., Melinder, A., Stridbeck, U., & Raja, A. Q. (2010). Beliefs about factors affecting the reliability of eyewitness testimony: A comparison of judges, jurors and the general public. *Applied Cognitive Psychology*, *24*(1), 122–133. https://doi.org/10.1002/ACP.1550

Magnussen, S., Wise, R. A., Raja, A. Q., Safer, M. A., Pawlenko, N., & Stridbeck, U. (2009). What judges know about eyewitness testimony: A comparison of Norwegian and US judges. *Http://Dx.Doi.Org/10.1080/10683160701580099*, *14*(3), 177–188. https://doi.org/10.1080/10683160701580099

Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, *11*(2), 103–121. https://doi.org/10.1007/S10339-009-0337-0/FIGURES/8

Mayer, R. (1992). *Thinking, problem solving, cognition*. W. H. Freeman.

Meissner, P., & Wulf, T. (2017). The effect of cognitive diversity on the illusion of control bias in strategic decisions: An experimental investigation. *European Management Journal*, *35*(4), 430–439. https://doi.org/10.1016/J.EMJ.2016.12.004

Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Lyons, P. M., & Otto, R. K. (2007). *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*. *930*, 95–0.

https://books.google.com/books/about/Psychological_Evaluations_for_the_Courts.html?hl=it&id=OsfI9dAVFvcC

Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial Allegiance among Expert Witnesses. *Http://Dx.Doi.Org/10.1146/Annurev-Lawsocsci-120814-121714*, *11*, 37–55. https://doi.org/10.1146/ANNUREV-LAWSOCSCI-120814-121714

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are Forensic Experts Biased by the Side That Retained Them? *Psychological Science*, *24*(10), 1889–1897. https://doi.org/10.1177/0956797613481812

Murrie, D. C., & Warren, J. I. (2005). Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Professional Psychology: Research and Practice*, *36*(5), 519–524. https://doi.org/10.1037/0735-7028.36.5.519

Nakhaeizadeh-, S., Dror, I., & Morgan, R. M. (2015). The emergence of cognitive bias in forensic science and criminal investigations. In *British Journal of American Legal Studies* (Vol. 527, Issue 2). https://heinonline.org/HOL/Page?handle=hein.journals/bjamles4&id=544&div=24&collection=journals

Neal, T. M. S. (2016). Are forensic experts already biased before adversarial legal parties hire them? *PLoS ONE*, *11*(4). https://doi.org/10.1371/JOURNAL.PONE.0154434

Neal, T. M. S. (2018). Discerning bias in forensic psychological reports in insanity cases. *Behavioral Sciences & the Law*, *36*(3), 325–338. https://doi.org/10.1002/BSL.2346

Neal, T. M. S., & Grisso, T. (2014). Assessment Practices and Expert Judgment Methods in Forensic Psychology and Psychiatry: An International Snapshot. *Criminal Justice and Behavior*, *41*(12), 1406–1421. https://doi.org/10.1177/0093854814548449

Neil Gowensmith, W., Murrie, D. C., Boccaccini, M. T., & McNichols, B. J. (2017). Field reliability influences field validity: Risk assessments of individuals found not guilty by reason of insanity. *Psychological Assessment*, *29*(6), 786–794. https://doi.org/10.1037/PAS0000376

Noppeney, U. (2021). Perceptual Inference, Learning, and Attention in a Multisensory World. In *Annual Review of Neuroscience* (Vol. 44, pp. 449–473). Annual Reviews. https://doi.org/10.1146/annurev-neuro-100120-085519

Ogloff, J. R. P., Roberts, C. F., & Roesch, R. (1993). The insanity defense: Legal standards and clinical assessment. *Applied and Preventive Psychology*, *2*(3), 163–178. https://doi.org/10.1016/S0962-1849(05)80122-2

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/J.TICS.2008.02.014

Packer, I. (2009). *Evaluation of criminal responsibility*. Oxford University Press.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). *Eyewitness Recall and Testimony.* https://psycnet.apa.org/record/2005-07316-004

Perez, O. (2014). *Can experts be trusted and what can be done about it? Insights from the biases and heuristics literature*. https://doi.org/10.13140/2.1.4839.0409

Peterson, J. L., Hickman, M. J., Strom, K. J., & Johnson, D. J. (2013). Effect of Forensic Evidence on Criminal Justice Case Processing. *Journal of Forensic Sciences*, *58*(SUPPL. 1), S78–S90. https://doi.org/10.1111/1556-4029.12020

Phillips, J. K. G., & Woodman, R. E. (2007). The Insanity of the Mens Rea Model: Due Process and the Abolition of the Insanity Defense. *Pace Law Review*, *28*. https://heinonline.org/HOL/Page?handle=hein.journals/pace28&id=463&div=29&collection=journals

Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K., & Grafman, J. (2000). Neural correlates of imaginal aggressive behavior assessed by positron emission tomography in healthy subjects. *American Journal of Psychiatry*, *157*(11), 1772–1781. https://doi.org/10.1176/APPI.AJP.157.11.1772

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience 2013 16:9*, *16*(9), 1170–1178. https://doi.org/10.1038/nn.3495

Prezenski, S., Brechmann, A., Wolff, S., & Russwinkel, N. (2017). A cognitive modeling approach to strategy formation in dynamic decision making. *Frontiers in Psychology*, *8*(AUG), 1335. https://doi.org/10.3389/FPSYG.2017.01335/XML/NLM

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *The Behavioral and Brain Sciences*, *22*(3), 341–365. https://doi.org/10.1017/S0140525X99002022

Qualtrics. (2022). *Qualtrics XM // The Leading Experience Management Software*. https://www.qualtrics.com/uk/?rid=ip&prevsite=en&newsite=uk&geo=IT&geomatch=uk

Rafter, N. (2008). The criminal brain: Understanding biological theories of crime. In *The Criminal Brain: Understanding Biological Theories of Crime*. https://doi.org/10.1176/appi.ps.60.9.1277-a

Raine, A. (2002). The Biological Basis of Crime. *Crime: Public Policies for Crime Control*.

Romero-Martínez, Á., González, M., Lila, M., Gracia, E., Martí-Bonmatí, L., Alberich-Bayarri, Á., Maldonado-Puig, R., Ten-Esteve, A., & Moya-Albiol, L. (2019). The brain resting-state functional connectivity underlying violence proneness: Is it a reliable marker for neurocriminology? A systematic review. *Behavioral Sciences*, *9*(1). https://doi.org/10.3390/BS9010011

Saferstein, R. (2018). Criminalistics an Introduction to Forensic Science. In *Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan di RSUD Kota Semarang* (Vol. 3).

Sajous-Turner, A., Anderson, N. E., Widdows, M., Nyalakanti, P., Harenski, K., Harenski, C., Koenigs, M., Decety, J., & Kiehl, K. A. (2020). Aberrant brain gray matter in murderers. *Brain Imaging and Behavior*, *14*(5), 2050–2061. https://doi.org/10.1007/S11682-019-00155-Y

Sartori, G., Scarpazza, C., Codognotto, S., & Pietrini, P. (2016). An unusual case of acquired pedophilic behavior following compression of orbitofrontal cortex and hypothalamus by a Clivus Chordoma. *Journal of Neurology*, *263*(7), 1454–1455. https://doi.org/10.1007/S00415-016-8143-Y

Scarpazza, C., Ferracuti, S., Miolla, A., & Sartori, G. (2018). The charm of structural neuroimaging in insanity evaluations: guidelines to avoid misinterpretation of the findings. *Translational Psychiatry*, *8*, 227. https://doi.org/10.1038/s41398-018-0274-8

Scarpazza, C., Miolla, A., Zampieri, I., Melis, G., Sartori, G., Ferracuti, S., & Pietrini, P. (2021). Translational Application of a Neuro-Scientific Multi-Modal Approach Into Forensic Psychiatric Evaluation: Why and How? *Frontiers in Psychiatry*, *12*, 70. https://doi.org/10.3389/FPSYT.2021.597918/BIBTEX

Scarpazza, C., Zampieri, I., Miolla, A., Melis, G., Pietrini, P., & Sartori, G. (2021). A multidisciplinary approach to insanity assessment as a way to reduce cognitive biases. *Forensic Science International*, *319*. https://doi.org/10.1016/J.FORSCIINT.2020.110652

Schneider, B., Chevallier, C., Dominguez, A., Bruguier, C., Elandoy, C., Mangin, P., & Grabherr, S. (2012). The forensic radiographer: A new member in the medicolegal team. *American Journal of Forensic Medicine and Pathology*, *33*(1), 30–36. https://doi.org/10.1097/PAF.0B013E31820C6AA3

Schug, R. A., Yang, Y., Raine, A., Han, C., & Liu, J. (2010). Structural and psychosocial correlates of birth order anomalies in schizophrenia and homicide. *Journal of Nervous and Mental Disease*, *198*(12), 870–875. https://doi.org/10.1097/NMD.0B013E3181FE7280

Schwarz, N. (2010). Meaning in context: Metacognitive experiences Met acognit ive Experiences in Consumer Judgment and Decision Making. In *The mind in context* (pp. 105–125).

Schwiedrzik, C. M., Melloni, L., & Schurger, A. (2018). Mooney face stimuli for visual perception research. *PLOS ONE*, *13*(7), e0200106. https://doi.org/10.1371/JOURNAL.PONE.0200106

SPSS Software. IBM Corp. Released 2022. IBM SPSS Statistics for Windows, Version 29.0. Armonk, NY: IBM Corp

Sternberg, R., & Ben-Zeev, T. (2001). *Complex cognition: The psychology of human thought.* https://psycnet.apa.org/record/2001-00285-000

Sterzer, P., & Kleinschmidt, A. (2007). A neural for inference in ambiguity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(1), 323–328. https://doi.org/10.1073/PNAS.0609006104/SUPPL_FILE/09006FIG6.PDF

Stoel, R. D., Berger, C. E. H., Kerkhoff, W., Mattijssen, E. J. A. T., & Dror, I. E. (2016). Minimizing Contextual Bias in Forensic Casework. In *Forensic Science and the Administration of Justice: Critical Issues and Directions* (pp. 67–86). https://doi.org/10.4135/9781483368740.n5

Strickland, N. H. (2015). Quality assurance in radiology: peer review and peer feedback. *Clinical Radiology*, *70*(11), 1158–1164. https://doi.org/10.1016/J.CRAD.2015.06.091

Tallarida, R. J., & Murray, R. B. (1987). Chi-Square Test. *Manual of Pharmacologic Calculations*, 140–142. https://doi.org/10.1007/978-1-4612-4974-0_43

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/SCIENCE.185.4157.1124

Tversky, A., psychology, D. K.-C., & 1973, undefined. (1973). Availability: A heuristic for judging frequency and probability. *Elsevier*, *5*, 207–232. https://www.sciencedirect.com/science/article/pii/0010028573900339

Ulery, B. T., Hicklin, R. A., Buscaglia, J. A., & Roberts, M. A. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, *7*(3). https://doi.org/10.1371/JOURNAL.PONE.0032800

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*(4), 339–345. https://doi.org/10.1111/j.1467-9280.2006.01708.x

Wegner, D. M. (1994). Ironic Processes of Mental Control. *Psychological Review*, *101*(1), 34–52. https://doi.org/10.1037/0033-295X.101.1.34

Wells, G., Psychology, E. O.-A. review of, & 2003, undefined. (2003). Eyewitness testimony. *Annualreviews.Org*, *54*, 277–295. https://doi.org/10.1146/annurev.psych.54.101601.145028

Wertheimer, M. (1985). A Gestalt perspective on computer simulations of cognitive processes. *Computers in Human Behavior*, *1*(1), 19–33. https://doi.org/10.1016/0747-5632(85)90004-4

Wertheimer, M. (2010). *A Gestalt Perspective on the Psychology of Thinking*. 49–58. https://doi.org/10.1007/978-3-642-03129-8_4

Williams, W. H., Chitsabesan, P., Fazel, S., McMillan, T., Hughes, N., Parsonage, M., & Tonks, J. (2018). Traumatic brain injury: a potential cause of violent crime? *The Lancet. Psychiatry*, *5*(10), 836. https://doi.org/10.1016/S2215-0366(18)30062-2

Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology*, *18*(4), 427–443. https://doi.org/10.1002/ACP.993

Wise, R. A., Sartori, G., Magnussen, S., & Safer, M. A. (2014). An examination of the causes and solutions to eyewitness error. *Frontiers in Psychiatry*, *5*(AUG), 102. https://doi.org/10.3389/FPSYT.2014.00102/XML/NLM

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the Reliability of Eyewitness Memory. *Perspectives on Psychological Science*, *13*(3), 324–335. https://doi.org/10.1177/1745691617734878

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (2016). The Comprehensive System for the Rorschach: A Critical Examination: *Https://Doi.Org/10.1111/j.1467-9280.1996.Tb00658.x*, *7*(1), 3–10. https://doi.org/10.1111/J.1467-9280.1996.TB00658.X

Woody, R. H. (2016). Psychological Testimony and the Daubert Standard. *Psychological Injury and Law 2016 9:2*, *9*(2), 91–96. https://doi.org/10.1007/S12207-016-9255-5