Strategic choice in linear sequential unmasking☆

Roger Koppl

Department of Finance, Syracuse University, 721 University Avenue, Syracuse, NY 13244, USA

1. Introduction

Without forensic science, more innocent persons would be falsely convicted and more guilty persons would escape conviction. Without forensic science, in other words, there would be higher rates of Type I and Type II errors in the criminal justice system. Forensic science improves the criminal justice system by reducing these error rates. It is therefore desirable to employ all available tools to minimize such errors while, of course, recognizing the possibility of a tradeoff between them. At least one such tool has been underutilized: descriptive mathematical decision theory.

This study uses descriptive decision theory to evaluate crime lab protocols. By way of illustrative application, it asks whether confidence attributions would help to make linear sequential unmasking [1] an improvement on sequential unmasking [2]. The study is innovative because it puts forensic scientists into the model of forensic science. We need to be sure proposed protocols, procedures, and practices will have their intended effects. We cannot be sure of such effects, however, without including in our analysis a model of how the proposed process would be applied by forensic scientists who share the universal human tendency to respond to incentives.

No prior analysis has used descriptive decision theory to evaluate linear sequential unmasking (LSU). Doing so leads to an unexpected and potentially important conclusion about the role of confidence attributions in LSU. The consequences of confidence attributions depend on the organizational context, which influences the costs and benefits of different examiner choices.

Dror et al. [1] tentatively call for the use of confidence attributions to discourage forensic scientists from changing their determinations. The decision theoretic analysis of this paper will address only these confidence attributions. It does not address other mechanisms "to impose limits and restrictions for when examiners are permitted to revisit and alter their initial analysis of trace evidence" [1] or any other aspect of LSU. The analysis raises an important caution on one potential LSU mechanism: confidence attributions. And it illustrates an important approach to evaluating many issues in the organization of forensic science, including other potential LSU mechanisms. The descriptive decision theoretic analysis of this paper finds that one candidate for achieving the stated end of LSU depends sensitively upon an important element of the crime lab's operation and culture, an element which may be called the lab's "error tolerance regime." The value of any measure such as LSU is highly dependent on the organizational context, which influences the costs and benefits of different examiner choices. Confidence attributions may be informative or meaningless depending on the error tolerance regime. No proposal to reduce bias and error in forensic science should neglect the organizational determinants of the error tolerance regime.

This study illustrates the value of one approach to studying crime lab protocols that has so far been underutilized, descriptive decision theory. No suggestion is made, however, that other approaches are somehow inferior. We need to draw on all available sources of knowledge and insight. It seems plausible to suggest that the link here drawn between the error tolerance regime and the value of confidence attributions would have been hard to find with other techniques of analysis. Each technique will have its strengths and weaknesses, however, and we should use multiple techniques to uncover the truth. The value of descriptive decision theory may be greater when we are entering a relatively new area with correspondingly little empirical evidence. While the general phenomenon of blinding or unmasking has been studied quite a bit [3], less is known about how to apply the general principle of masking to the particular problems of forensic science. Lockhart and Satya-Murti are probably right to say of forensic science, "Research examining blinding approaches is still rather novel" [4].

Although this study employs mathematical decision theory, it is self-contained and uses only some basic algebra and probability theory. It draws a perhaps surprising conclusion about the importance of error tolerance regimes. It thereby suggests that the innovative way in which it uses decision theory may have value for examining and perhaps improving forensic science.

2. Methods

Mathematical decision theory is based on the somewhat humdrum insight that people generally try to maximize benefits net of costs as those costs and benefits are subjectively perceived. Sometimes costs and benefits are monetary, but often they are not. When choosing between a

bookkeeper's job at a grocery store and a bookkeeper's job at a pharmacy, the main issue may well be pay. Who pays more? When choosing between a job as bookkeeper and fuel tank cleaner, however, money may be a relatively minor consideration. The great discomfort associated with cleaning fuel tanks is a cost of such work that may be perceived (subjectively) as very high. The costs and benefits discussed below are subjectively perceived by forensic examiners whether or not they have anything to do with money.

Bernoulli [5] is commonly cited as the earliest example of mathematical decision theory of the sort described here. Bernoulli [6] is an English translation. His 1738 article attempted to resolve the St. Petersburg Paradox. Imagine a casino that offers to flip a fair coin until tails come up for the first time. As long as heads come, they will keep flipping. Once tails come up, they stop flipping. If tails come up for the first time on the first flip (T), the casino will pay $2.00. If tails come up for the first time on the second flip (HT), the casino will pay $4.00. If tails come up for the first time on the third flip (HHT), the casino will pay $8.00. And so on. The payoff is 2^n, where n is the number of the flip in which the coin first comes up tails. The probability of getting all heads until the nth flip is $\frac{1}{2^n}$. Thus, the expected value is $\sum_{n=1}^{\infty} \left(\$2^n \times \frac{1}{2^n} \right)$ or $\sum_{n=1}^{\infty} \$1$, which is infinite. The paradox is that no one would be willing to pay more than a relatively small sum to play this game even though it has an infinite expected value.

Bernoulli resolved the paradox by postulating that the utility (*emolumentum*) of a person's winnings rises with the logarithm of those winnings. With this assumed utility function, the expected *utility* of the gamble is finite, and a prospective gambler would be willing to pay only a relatively low price to play. Menger [7] subjected this problem to a penetrating examination. Menger [8] is an English translation. Menger showed that Bernoulli's solution did not wipe out all paradox. Given an unbounded utility function, one can always find a gamble with an infinite expected utility that, nevertheless, no one would pay more than a relatively small sum to play. In a book first published in 1944, von Neumann and Morgenstern [9] greatly refined Bernoulli's general approach in the context of developing the mathematical theory of games. Building on this book, Abraham Wald developed modern decision theory as the theory of solitary individuals playing "games" against "nature." Thus, as DeGroot [10] explains, the modern decision theory used in this paper was in "many respects... an outgrowth, and a special case, of the *theory of games*," and it "was developed mainly by Abraham Wald during the 1940s."

Decision theory has been applied to forensic science. Usually, Bayes' theorem is central to such applications, and the theory may be called "Bayesian decision theory." In the relatively narrow context of this paper, Bayes' theorem does not arise. In the broader context of decision theoretic forensic science, Bayes' theorem is an important tool. The earliest example of Bayesian decision theory in forensic science might be Lindley [11]. The Bayesian decision theoretic tradition in forensic science is associated, though not exclusively, with the School of Criminal Justice at the University of Lausanne. Taroni et al. [12] is an important book length treatment. Contributions to this tradition include Taroni et al. [12], Biedermann et al. [13], Gittelson et al. [14], Gittelson et al. [15], and Gittelson et al. [16]. The probabilities used in Bayesian decision theory are usually interpreted as subjective degrees of belief. Taroni et al. [12], for example, say, "*Bayesian epistemology* claims that the tools in the basket of scientific method, whatever they might be, must surely satisfy the standard of pragmatic coherence for combining and changing degrees of belief: therefore, the fundamental tool of scientific method is Bayes' theorem" (p. 43, emphasis in original).

Decision theory may be "descriptive" or "normative." Sabine [17] says, "The general division of sciences into descriptive and normative has long been one of the commonest devices used in classification." The difference between descriptive and normative is the difference between "is" and "ought." A theory is "descriptive" if it describes what is. A theory is "normative" if it tells us what we should do or how we should

do something. Decision theory in forensic science is usually considered normative. For example, Gittelson et al. [15] say, "Decision theory provides a normative model for making rational decisions under uncertainty." But decision theory can be used descriptively. Becker [18], for example, gives a decision theoretic account of crime. He "assumes that a person commits an offense if the expected utility to him exceeds the utility he could get by using his time and other resources at other activities." This expected utility depends on factors such as the "probability of conviction" and the "income available in legal activities." Becker's analysis is descriptive and not normative.

The normative use of decision theory in forensic science requires a scientifically appropriate model of the underlying phenomena. The normative use of Bayesian decision theory to guide forensic DNA analysis, for example, requires an appropriate model of the human genome. The forensic scientist's model of the human genome must imply that an individual is either homozygous or heterozygous at a given locus. A person has either one or two alleles at a given locus, not zero, three, or more than three.

In descriptive analysis, it is possible, though certainly not required, that the decision maker believes a false theory. Consider, for example, Lentini's [19] demonstration that crazed glass is not a reliable indicator of exposure to rapidly increasing temperature. This result excludes crazed glass from the *normative* use of decision theory in fire investigation. But in a *descriptive* decision theoretic analysis of fire investigation, the use of crazed glass as an arson indicator might be legitimate and important. We cannot understand how the fire investigation went wrong unless we recognize that the fire investigator thought crazed glass was a reliable indicator of exposure to rapidly increasing temperature.

The normative use of decision theory also requires that the forensic scientist's preferences be free of inappropriate elements. The scientist's utility function should not, for example, be influenced by a desire to vindicate the police theory of a case or the desire to abuse or oppress ethnic minorities. Normative analysis requires that the forensic scientist's preferences or utility be free of such inappropriate biases. Lindley [20] says that the forensic scientist's "utility" is "imposed by society." Taroni et al. [21] seem to take a more individualistic approach to utility. They favorably cite Lindley [22] saying, "The value[s] inserted for the utilities and probabilities are in no sense correct and any other values wrong. They represent the decision-maker's individual preferences and may be modified by him." (Lindley [22] as quoted in Taroni [21]). Citing Lindley [20], Gittelson et al. [15] say the "utility function is imposed by the objectives and preferences of society." Gittelson et al. [16], however, take a more individualistic approach. They say that peak-height thresholds in DNA analysis depend in part on "the scientist's objectives (i.e., preferences regarding false exclusions and false inclusions). These parameters vary in function of the case, the scientist and the laboratory." They say, "[T]he scientist's preferences will vary from one scientist to another, and may also vary from one case to another in function of the severity of the case."

The analysis of this paper is descriptive and not normative. It does not attempt to represent the problems forensic scientists should solve or how they should solve them. These normative questions are important. For that very reason, however, it is also important to ask whether a given proposal for laboratory protocols will or will not bring us closer to the normatively correct ideal. The following analysis addresses an important issue of descriptive analysis. Will LSU confidence attributions *in fact* improve forensic science relative to sequential unmasking (SU)? The following analysis suggests that a positive answer is contingent on the overarching organization of the crime laboratory in general and its error tolerance regime in particular. The treatment of and attitude toward error in a crime lab is important to determining the consequences of using confidence attributions to achieve the goals of linear sequential unmasking.

3. Results and discussion

It was noted earlier that forensic science serves an error correcting role within the criminal justice system. A kind of autonomy, therefore, is desirable for forensic science: To the extent possible, forensic science determinations should depend on scientific factors alone and not scientifically irrelevant factors such as the identity of the suspect or the nature of the suspected crime. When task-irrelevant, case specific information influences the scientific interpretation of evidence, the final disposition of the criminal case may deviate from the scientifically corrected outcome and toward the outcome that would occur without forensic science. To the extent that such deviations occur, forensic science is not performing its error-correction function. Masking procedures such as sequential unmasking [2] and linear sequential unmasking [1] have been recommended to minimize them. These measures are meant to reduce "observer effects."

Observer effects exist when extra-scientific considerations influence a forensic examiner's scientific judgment, if only unconsciously. Observer effects are biases and errors of observation caused by the state of the observer. Forensic scientists observe evidence and make decisions about what they have observed. Those decisions may be biased or erroneous if the observations are biased or erroneous. The state of the observer may induce errors and biases in the observation. In such a case, the state of the observer is the *cause* and the error or bias in observation is the *effect*. Thus, the bias or error in observation is the *effect* of the state of the *observer*, and we may speak of *observer effects*.

The distinction between observation and decision has been emphasized quite strongly by Dror [23] who explains and documents this distinction in great detail. His "Hierarchy of Expert Performance" (HEP) ramifies the distinction into four aspects of expert performance relating to observations and four aspects of expert performance relating to conclusions.

As Koppl et al. [24] note, the "three keys to observer effects in forensic science" are "the analyst's state of *expectation*, the analyst's state of *desire*, and the degree of *ambiguity* in the evidence being examined." In this paper, the degree of ambiguity of the evidence is given outside the analysis. The paper thus focuses on the analyst's state of expectation and, especially, desire. Forensic examiners make choices and those choices are influenced by desires. In decision theory, as we have seen, desires are somewhat blandly labeled "preferences" or "utility." The theory examines how a person with given preferences (or a given "utility function") chooses when the outcome of that choice depends on contingencies that the person cannot control.

Krane et al. [2] proposed sequential unmasking as "a means of minimizing observer effects" in forensic science. Sequential unmasking is an administrative control process similar to that used in double-blind research studies. It prevents forensic scientists from learning potentially biasing information until after they have made the decisions that might be so biased. Some potentially biasing information, including the reference sample, must be observed to complete a forensic analysis. Krane et al. [2] suggest that potentially biasing, but task-relevant information be *sequenced* such that potentially biasing information is not released to the forensic examiner until after they have made the decisions that might be biased by that information. A case manager who is privy to all case information sequences information for the examiner. Importantly, Krane et al. [2] propose, "sequencing the laboratory workflow such that evidentiary samples are interpreted, and the interpretation is fully documented, before reference samples are compared." In the case of DNA typing, forensic examiners should "call the alleles" before seeing the reference sample.

At each stage, the examiner could change earlier determinations as long as all such changes are fully documented. (This possibility is assumed but not stated in Krane et al. [2].) Each bit of task-relevant, but potentially biasing information is initially masked from the examiner. In *sequential* fashion, however, such information is *unmasked*, until all task-relevant information is available to the examiner, and this procedure is called *sequential unmasking*.

Although sequential unmasking applies broadly, it was originally proposed in the context of forensic DNA testing, which "is not immune to bias, especially when subjective interpretation is involved" [25]. Jeanguenat, Budowle, and Dror [25] have recently developed an important account of the surprising variety of ways in which cognitive bias may enter DNA analysis. They propose a suite of measures to combat such bias including, but not restricted to masking protocols such as sequential unmasking and linear sequential unmasking.

Krane et al. [2] explain that unwanted observer effects are most likely "when the underlying data are ambiguous and the scientist is exposed to domain-irrelevant information." Dror et al. [1] address this role of ambiguity in creating observer effects by proposing "linear sequential unmasking" (LSU) as a refinement of sequential unmasking. At each stage in the sequential unmasking process, examiners would be required to "specify their confidence" in each of the determinations made. Somewhat tentatively, Dror et al. [1] propose a color system with "'green for certainty,' 'yellow for low certainty,' and 'red for questionable.'" This procedure is meant to constrain examiners' ability to revise their earlier determinations, thereby creating a better "balance between restrictive procedures that forbid analysts from changing their opinion and those that allow unlimited and unrestricted changes."

Dror et al. [1] explain that sequential unmasking and linear sequential unmasking are both "linear in the sense that one must begin with the trace evidence before being exposed to and working with the reference material, thus working from the evidence to the suspect, rather than from the suspect to the evidence." One may conjecture that *linear* sequential unmasking is more "linear" than sequential unmasking because it is intended to discourage changes to earlier determinations, and such changes might invoke the idea of a spiral rather than a straight line.

Dror et al. [1] seem to imagine that forensic examiners will apply the confidence attributions green, yellow, and red without calculating the costs and benefits of different confidence attributions. Everyday experience, however, confirms that one's confidence in a judgment depends on the consequences of that judgment being wrong. For example, you may have more confidence that a rickety table can support a bag of dog food than an infant child of the same weight. A forensic examiner's confidence in a judgment will typically depend in part on the consequences of that judgment being reversed later in the process of linear sequential unmasking. But if that is so, then LSU confidence attributions must be viewed as variables subject to *strategic choice* by forensic examiners. They are *strategic variables*. Therefore, it is both possible and necessary to subject the choice of confidence attribution to a decision-theoretic analysis.

Dror et al. [1] say, "Clearly, there is a difference between revisiting and revising prior analysis decisions that an examiner initially perceived as only suggestive (low confidence) versus revisiting and revising analysis decisions that an examiner perceived as obvious and certain (high confidence)." Presumably, then, the examiner experiences the need for revision as a (subjective) cost, and the cost is higher the greater the confidence attribution. If so, then (by virtue of the logical symmetry of costs and benefits) it is also a subjective *benefit* to the examiner if a determination of theirs does *not* require revision. It is a benefit if it is supported by subsequently unmasked information. And this benefit is higher the greater the confidence attribution. We should therefore recognize that confidence attributions must be *chosen* by the examiner and are, therefore, *strategic variables* whose value will be chosen as the outcome of a cost-benefit calculation.

After looking at the crime-scene sample the examiner will estimate, if only implicitly, the probabilities of the different statements they might make about the sample. Let us consider one such statement: "The peak labeled '12' at locus 'D3S1358' is an artifact." Let p be the subjective probability that the statement will be reversed at the end of the analysis. Thus, p is what Koppl et al. [24] call "reversal risk." It might be expected that the examiner would include the statement in their

Table 1

An illustrative payoff matrix.

	Subsequently unmasked information tends to support statement	Subsequently unmasked information tends to undermine statement
Green	3	−6
Yellow	2	−4
Red	1	−2

The rows represent confidence levels. The examiner must pick one of the rows. The columns represent the tendency of the subsequently unmasked information to support or undermine the examiner's statement. Each box contains a number showing the cost or benefit to the examiner of the subsequently unmasked information given the chosen level of confidence associated with the statement.

Table 2

Choosing the optimal statement.

	$0 \leq p < \frac{1}{3}$	$\frac{1}{3} < p < \frac{2}{3}$	$\frac{2}{3} < p \leq 1$
Determine that peak is an artifact	Choose confidence attribution Green and receive the payoff $3 − 9p$, which ranges between 0 and 3 for these p values.	Choose confidence attribution Red and receive the payoff $1 − 3p$, which ranges between −1 and 0 for these p values.	Choose confidence attribution Red and receive the payoff $1 − 3p$, which ranges between −2 and −1 for these p values.
Determine that it cannot be decided whether the peak is an artifact	No confidence attribution can be made in this case. Receive the payoff 0.	No confidence attribution can be made in this case. Receive the payoff 0.	No confidence attribution can be made in this case. Receive the payoff 0.
Determine that peak is *not* an artifact	Choose confidence attribution Red and receive the payoff $3p − 2$, which ranges between −2 and −1 for these p values.	Choose confidence attribution Red and receive the payoff $3p − 2$, which ranges between −1 and 0 for these p values.	Choose confidence attribution Green and receive the payoff $9p − 6$, which ranges between 0 and 3 for these p values.

The rows represent alternative determinations about the peak at locus D3S1358. The examiner must pick one of the rows. The columns represent ranges of p, the examiner's subjective reversal risk for the determination that the peak is an artifact. The subjective reversal risk for the determination that the peak is *not* an artifact is $1 − p$. Each box indicates 1) what confidence attribution (if any) would maximize the examiner's payoff, 2) what that payoff would be as a function of p and, finally, 3) the range of possible payoffs for that contingency. If $0 \leq p < \frac{1}{3}$, the examiner will determine that the peak is an artifact and choose confidence attribution Green for that determination. If $\frac{1}{3} < p < \frac{2}{3}$, the examiner will determine that it cannot be decided whether the peak is an artifact. If $\frac{2}{3} < p \leq 1$, the examiner will determine that the peak is *not* an artifact and choose confidence attribution Green for that determination.

Table 3

A more general payoff matrix.

	Subsequently unmasked information tends to support statement	Subsequently unmasked information tends to undermine statement
Green	3	$−3\gamma$
Yellow	2	$−2\upsilon$
Red	1	$−\rho$

The rows represent confidence levels. The examiner must pick one of the rows. The columns represent the tendency of the subsequently unmasked information to support or undermine the examiner's statement. Each box contains a number or expression showing the cost or benefit to the examiner of the subsequently unmasked information given the chosen level of confidence associated with the statement.

initial analysis if the reversal risk is less than one half, $p < 0.5$. As we shall see, however, the threshold value for p may well be < 0.5 and in some cases, substantially less.

If the forensic scientist is to make the statement affirming an artifact at locus D3S1358, they must also *choose* a confidence attribution (Green, Yellow, Red) to associate with it. After making the statement and picking a confidence attribution they will be exposed to some contextual information, such as the reference sample. That context may tend to support the statement or it may tend to undermine the statement. Thus, if the examiner makes the statement, they would face a payoff matrix such as the one shown in Table 1. The assumed costs and benefits will be modified later to provide a more general analysis. But it is helpful to first review a relatively simple example that produces a surprising result. The simple example is merely illustrative. It shows that costs and benefits influence both the determinations made and the confidence attributions associated with those determinations. This exercise sets the stage for the more careful and, perhaps, realistic analysis that follows. This more complete analysis points to the importance of organizational design in establishing a desirable error tolerance regime.

Table 1 reflects the principles about examiner preferences, which, as we have seen, were invoked by Dror et al. [1]. It is better for the examiner (it is a subjective benefit) if their statement is supported by the subsequently unmasked context information. That perceived benefit is greater the higher the level of confidence the examiner attaches to the

determination. The forensic examiner's statement may, however, be undermined by the subsequently unmasked context information. They would experience that undermining as a subjective cost. That perceived cost is greater the higher the level of confidence the examiner attached to the determination.

The examiner will choose the confidence attribution that maximizes their expected payoff. The payoff to choosing Green will be 3 if the statement holds up, and −6 if it reversed. The subjectively perceived reversal risk is p and, therefore, the examiner's subjective probability that the statement will not be reversed is $1 − p$. Thus, $3(1 − p)$ is the prospective benefit of choosing confidence attribution Green and $6p$ is the prospective cost. The net prospective benefit of choosing confidence attribute Green is therefore $3(1 − p) − 6p$, which is $3 − 9p$. Similar calculations can be made for Yellow and Red. The expected net benefits are:

Green: $3 − 9p$

Yellow: $2 − 6p$

Red: $1 − 3p$

If $p < \frac{1}{3}$, Green is the confidence attribution with the highest net benefit. If $p = \frac{1}{3}$, the three confidence attributions all have the same net benefit, which is 0. If $p > \frac{1}{3}$, Red is the confidence attribution with the

highest net benefit. From the analysis so far, it seems that the examiner will rarely or never pick confidence level Yellow.

From the analysis so far, we may conclude that *if* the examiner chooses to make the statement that the peak at D3S1358 is an artifact, *then* they will choose confidence attribution Green if $p < \frac{1}{3}$ and Red if $p > \frac{1}{3}$. The forensic examiner will choose Green if they perceive the reversal risk to be low, and Red if they perceive reversal risk to be high. But will they make that statement? They can also state that the peak is not an artifact or that it is not possible to determine whether it is an artifact. To complete our analysis we need to consider the costs and benefits of making each of these statements.

Consider the statement that the peak is not an artifact. The case here is perfectly parallel to the one considered above. Thus, *if* the examiner chooses to make the statement that the peak at D3S1358 is *not* an artifact, *then* she will choose confidence attribution Green if the reversal risk, q, for *that* statement is low, and she will choose confidence attribution Red if it is high. The expected net benefits are:

Green: $3 - 9q$

Yellow: $2 - 6q$

Red: $1 - 3q$

But $q = 1 - p$. When expressed in terms of p, therefore, the expected net benefits are:

Green: $9p - 6$

Yellow: $6p - 4$

Red: $3p - 2$

Thus, *if* the examiner chooses to make the statement that the peak at D3S1358 is *not* an artifact, *then* they will choose confidence attribution Green if $p > \frac{2}{3}$ and Red if $p < \frac{2}{3}$.

If the examiner states that it is not possible to determine whether the peak is an artifact, then their determination cannot be sustained or reversed. For this determination, therefore, it is impossible to acquire the benefit of the statement being supported by subsequent analysis, and it is impossible to suffer the cost of the statement being reversed by subsequent analysis. Thus, the payoff for this statement is 0.

We can now determine which statement has the highest value to the examiner. Table 2 shows the payoff to each statement as a function of p, the subjectively perceived reversal risk for the determination that the peak at D3S1358 is an artifact. The examiner can always ensure a payoff of at least 0 by making the determination that it cannot be decided whether the peak is an allele. If p is low enough, however, the examiner will determine that the peak is an artifact and choose confidence attribution Green for that determination. If p is high enough, the examiner will determine that the peak is *not* an artifact and choose confidence attribution Green for that determination. The examiner will never choose the confidence attributions Yellow and Red.

If LSU created the costs and benefits for forensic examiners given in Table 1, we would only rarely find examiners choosing confidence attributions other than Green. While this result is merely illustrative, it shows that in some circumstances the confidence attributions of LSU could be meaningless. As we shall see, the reality could be very far from this result or, instead, close to it, depending on the error tolerance regime.

The analysis so far has taken examiner preferences as given. Examiner preferences depend, however, on the overall organization of the crime lab. In sequential unmasking and linear sequential unmasking, the examiner is asked to characterize the unknown sample before seeing the reference sample. Information revealed subsequently may induce the examiner to change their determinations. Such a change could be construed as showing that the earlier determination was an error. How will such an "error" be treated within the lab? Will it be viewed as normal or a "monster" in the sense of Lakatos [26]?

James Reason [27] has argued, "It is now widely held among human reliability specialists that the most productive strategy for dealing with active errors is to focus upon controlling their consequences rather than upon striving for their elimination." Reason suggests making organizations "error tolerant" in part by building in "redundancies." As Reason uses the term, it is the organization or system that is error tolerant, not any individual. But an organization or system cannot be error tolerant if those who make errors are afraid of being caught in an error from fear of blame, scorn, dismissal or the like.

In an error-tolerant organization, individuals are not subject to blame and opprobrium when errors are discovered. Rather, they are taught how to cope with errors to minimize their harmful consequences. As Rouse and Morris [28] have put it, "Errors are *not* inherently unacceptable; however, it may be that the consequences of error are unacceptable. From this perspective, one may be able to tolerate errors as long as consequences can be controlled."

Not all organizations, however, are error tolerant. Indeed, it seems fair to say the most crime labs in the US are error *in*tolerant. Blame becomes an important category in error intolerant organizations. Pyrek [29] decries the "destructive, fatalistic blame game" in forensic science."

A crime lab, then, can create an error tolerant or an error intolerant environment, which we may call the "error tolerance regime." And the error tolerance regime largely determines the examiner's subjective costs and benefits of making different determinations and choosing different confidence attributions to associate with those determinations. Linear sequential unmasking relies on the common-sense intuition that it will seem worse to revise a determination the higher the confidence attribution associated with that determination. The error tolerance regime influences how steep the curve is. It influences how rapidly the costs of reversal rise with confidence attribution.

To achieve a more general analysis than that given above, three new variables are introduced to reflect the error tolerance regime. These variables can be adjusted to reflect different assumptions about how rapidly the cost of reversal rises with confidence attribution. Table 3 shows the payoff matrix with these variables included. Assume $0 \le \rho \le \nu \le \gamma$. If $\rho = \nu = \gamma = 2$, then Table 3 reduces to Table 1. In that case, the costs of reversal rise linearly with confidence attributions, and confidence attributions become meaningless. If costs rise more than linearly, however, confidence attributions may be more meaningful. Consider the case in which $\rho = 0$ and $\nu = \gamma = 2$. In this case, the examiner can always guarantee themself a payoff of $1 - p$ by choosing confidence attribution Red. In this case, they will choose Red if $\frac{1}{4} < p < \frac{1}{2}$ and Green if $p < \frac{1}{4}$.

With the particular costs of reversal used in the last example, Yellow is never chosen. Unfortunately, we cannot simply choose the three reversal-cost variables. If we could, then we could engineer the values of p for which each confidence attribution would be chosen. We could shrink the range for Green to any arbitrarily small interval by picking a sufficiently large value for γ. It can be shown that Yellow will be chosen over Red if $p < \frac{1}{1 + 2\nu - \rho}$, which we could make as big or small as we want by picking ρ and ν appropriately. To enlarge the range for Red, decrease ρ or increase ν. To enlarge the range for Yellow, increase ρ or decrease ν. To see this, let $\nu = \rho$. In this case, Yellow will be chosen over Red if $p < \frac{1}{1 + \nu}$. Red will never be chosen if we set $\nu < 1$. As we increase ν, the value of p that pushes us into the Yellow range shrinks to as small an interval as we may wish. In general, the more steeply the costs of reversal rise with confidence attribution, the greater the incentive to choose a lower confidence attribution.

The consequences of confidence attributions in linear sequential unmasking depend on the error tolerance regime. If the error tolerance regime causes the costs of reversal to rise steeply, confidence attributions will tend to be low. If it causes them to rise slowly, confidence attributions will tend to be high. In all cases, confidence attributions are strategic variables. In extreme cases, confidence attributions are

meaningless. It is not sufficient, therefore, to include confidence attributions in a crime lab's LSU protocols. Crime labs must also be error tolerant organizations with well-designed error tolerance regimes. Unfortunately, it is not clear how that is to be done. Further research is required, then, to learn how to create desirable error tolerance regimes in modern crime laboratories.

Using descriptive decision theory to analyze linear sequential unmasking has led us to unexpected conclusions that might be difficult or impossible to reach by other methods. It thus seems plausible to suggest that there may be other, perhaps many other, fruitful applications of descriptive decision theory in forensic science.

References

[1] I. Dror, W. Thompson, C. Meissner, I. Kornfield, D. Krane, M. Saks, et al., Context management toolbox: a linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision-making, J. Forensic Sci. 60 (2015) 1111–1112.

[2] D. Krane, S. Ford, J. Gilder, K. Inman, A. Jamieson, R. Koppl, et al., Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation, J. Forensic Sci. 53 (2008) 1006–1007.

[3] Scott H. Podolsky, David S. Jones, Ted J. Kaptchuk, From trials to trials: blinding, medicine, and honest adjudication, in: Christopher T. Robertson, Aaron S. Kesselheim (Eds.), Blinding as a Solution to Bias, Elsevier Academic Press, Amsterdam, 2016, pp. 45–58.

[4] J.J. Lockhart, S. Satya-Murti, Diagnosing crime and diagnosing disease: bias reduction strategies in the forensic and clinical sciences, J. Forensic Sci. 62 (2017) 1534–1541.

[5] D. Bernoulli, Specimen Theoriae Novae de Mensura Sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae, 5 (1738), pp. 175–192.

[6] D. Bernoulli, Exposition of a new theory on the measurement of risk, Econometrica 22 (1954) 23–36.

[7] K. Menger, Das Unsicherheitsmoment in der Wertlehre Betrachtungen im Anschluß an das sogenannte Petersburger Spiel, Z. National. 5 (1934) 459–485.

[8] K. Menger, The role of uncertainty in economics, in: K. Menger (Ed.), Selected Papers in Logic and Foundations, Didactics, Economics, D. Reidel Publishing Company, Dordrecht, Boston, and London, 1979, pp. 259–278.

[9] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, 3rd ed, Princeton University Press, Princeton, 1953.

[10] M.H. Degroot, Decision theory, in: Balakrishnan (Ed.), Methods and Applications of Statistics in Business, Finance, and Management Science, John Wiley & Sons, Inc, Hoboken, 2010, pp. 122–132.

[11] D.V. Lindley, A problem in forensic science, Biometrika 64 (1977) 207–213.

[12] F. Taroni, S. Bozza, A. Bidermann, P. Garbolino, C. Aitken, Data Analysis in Forensic Science: A Bayesian Decision Perspective, John Wiley and Sons, Ltd, West Sussex, 2010.

[13] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, Forensic Sci. Int. 177 (2008) 120–132.

[14] S. Gittelson, A. Biedermann, S. Bozza, F. Taroni, The database search problem: a question of rational decision making, Forensic Sci. Int. 222 (2012) 186–199.

[15] S. Gittelson, S. Bozza, A. Biedermann, F. Taroni, Decision-theoretic reflections on processing a fingermark, Forensic Sci. Int. 226 (2013) e42–e47.

[16] S. Gittelson, A. Biedermann, S. Bozza, F. Taroni, Decision analysis for the genotype designation in *low-template-DNA* profiles, Forensic Sci. Int. Genet. 9 (2014) 118–133.

[17] G. Sabin, Descriptive and normative sciences, Philos. Rev. 21 (1912) 433–450.

[18] G. Becker, Crime and punishment: an economic approach, J. Polit. Econ. 76 (1968) 169–217.

[19] J. Lentini, Behavior of glass at elevated temperatures, J. Forensic Sci. 37 (1992) 1358–1362.

[20] D.V. Lindley, Probability and the law, J. R. Stat. Soc. Ser. D (The Statistician) 26 (1977) 203–220.

[21] F. Taroni, S. Bozza, C. Aitken, Decision analysis in forensic science, J. Forensic Sci. 50 (2004) 1–12.

[22] D.V. Lindley, Making Decisions. 2nd Rev.ed, John Wiley & Sons, London, 1985.

[23] I.E. Dror, A hierarchy of expert performance, J. Appl. Res. Mem. Cogn. 5 (2016) 121–127.

[24] R. Koppl, D. Charlton, I. Kornfield, D. Krane, M. Risinger, C. Robertson, et al., Do observer effects matter? A comment on Langenburg, Bochet, and Ford, Forensic Sci. Policy Manag. 6 (2015) 1–6.

[25] A.M. Jeanguenat, B. Budowle, I.E. Dror, Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias, Sci. Justice 57 (2017) 415–420.

[26] Imre Lakatos, Proofs and Refutations: The Logic of Mathematical Discovery, Cambridge University Press, Cambridge, 1976.

[27] J. Reason, Human Error, Cambridge University Press, Cambridge and New York, 1990.

[28] W. Rouse, N. Morris, Conceptual design of a human tolerant interface for complex engineering systems, Automatica 23 (1987) 231–235.

[29] K.M. Pyrek, Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Death Investigation System, Academic Press, Boston, 2007.