



The database search problem: A question of rational decision making

S. Gittelsohn^{a,*}, A. Biedermann^a, S. Bozza^b, F. Taroni^a

^a University of Lausanne, Ecole des Sciences Criminelles, Institut de Police Scientifique, le Batochime, 1015 Lausanne, Switzerland

^b University Ca' Foscari, Dipartimento di Economia, San Giobbe-Cannaregio 873, 30121 Venice, Italy

ARTICLE INFO

Article history:

Received 24 October 2011

Received in revised form 20 March 2012

Accepted 28 May 2012

Available online 3 July 2012

Keywords:

Database search

Evidential value

Bayesian decision theory

Influence diagrams

ABSTRACT

This paper applies probability and decision theory in the graphical interface of an influence diagram to study the formal requirements of rationality which justify the individualization of a person found through a database search. The decision-theoretic part of the analysis studies the parameters that a rational decision maker would use to individualize the selected person. The modeling part (in the form of an influence diagram) clarifies the relationships between this decision and the ingredients that make up the database search problem, i.e., the results of the database search and the different pairs of propositions describing whether an individual is at the source of the crime stain. These analyses evaluate the desirability associated with the decision of 'individualizing' (and 'not individualizing'). They point out that this decision is a function of (i) the probability that the individual in question is, in fact, at the source of the crime stain (i.e., the state of nature), and (ii) the decision maker's preferences among the possible consequences of the decision (i.e., the decision maker's loss function). We discuss the relevance and argumentative implications of these insights with respect to recent comments in specialized literature, which suggest points of view that are opposed to the results of our study.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The 'classical' database search problem, as it is known throughout forensic and legal theory and practice, relates to a question of the following kind: "What is the strength of the evidence against a given individual found through a database search, when that individual is the only person in the database who presents the same analytical characteristics (such as a DNA profile) as those observed on a crime stain?" After intense and controversial debates, starting in the mid-1990s, and which seemed to have been settled during the last decade, the database search problem has once more become the object of several publications [1–4]. In particular, Schneider et al. [1] and Fimmers et al. [3] recently claimed that a single matching profile found through a database search reduces the evidential value of this match compared to a match found by other investigational means (i.e., a situation in which no database search was conducted). However, there are now ample counterarguments demonstrating that this is a misconception [e.g., 5–10] dating back to the NRC reports [11,12] and writings by Stockmarr [13]. These latter accounts are constructed around a conceptually unsuitable pair of propositions, defined as follows:

H_{db} : the source of the crime stain is in the database;

$\neg H_{db}$: the source of the crime stain is not in the database.

This contradicts probabilistic arguments that demonstrate an increase in the evidential value of a single database match when one considers the conventional and procedurally appropriate pair of propositions, which take the form of:

H_i : the crime stain comes from individual i ;

$\neg H_i$: the crime stain comes from someone else unrelated to individual i ;

where $i = 1, \dots, N$, and N is the size of the population of individuals who could have been at the source of the crime stain. The increase for this pair of propositions is due to the exclusion of $n - 1$ non-matching profiles (where n denotes the size of the database searched). This argument is now covered to a great extent in existing literature [e.g., 5–10,14,15], and currently appears to accumulate the most widespread support.

In their recent publication, Fimmers et al. [3] seek to take their argument in support of a decrease in the value of a database match a step further: they addressed the act of convicting a suspect and the probability that this conviction is false. That is, they passed from a purely probabilistic discourse to an argument invoking the act of choosing a particular option among several possible options. Their argumentation consists of a hypothetical case, in which investigators search for the individual at the source of a biological stain recovered on a crime scene. The investigators in this case consider a population of 100 million individuals ($N = 10^8$) as the

* Corresponding author. Tel.: +41 021 692 46 07; fax: +41 021 692 46 05.
E-mail address: simone.gittelsohn@unil.ch (S. Gittelsohn).

population of potential sources, and possess a database containing the profiles of one million of these individuals ($n = 10^6$). In this population, the DNA profile of the crime stain has a match probability of $\gamma = 10^{-6}$. However, in their example, Fimmers et al. [3] assume that the true source of the crime stain is *not* in the population considered by the investigators, that is, not among the $N = 10^8$ individuals, and consequently, not among the $n = 10^6$ profiles in the database (since the database contains profiles taken from the population of the $N = 10^8$ individuals). Assuming that “a suspect will certainly be convicted in every case in which the rarity of the corresponding DNA profile is [at least] one in a million” [3, p. 4],¹ that is, when $\gamma \leq 10^{-6}$, they then compare the probability of a false conviction in a *probable cause case*² to the probability of a false conviction in a *database search case*³ for an incriminating profile with a match probability of 10^{-6} :

- *Probable cause case*: “There is a suspect. The DNA profile of that person is determined and found to correspond to that of the crime stain. The person is going to be convicted on the sole basis of this correspondence. Given the assumptions in this example, we know that the conviction is erroneous, because the true author has escaped. How high are the odds, in our scenario, of this to happen by chance? The probability of the DNA profile is 1:1,000,000 and this is the probability for a correspondence by chance with the stain. The probability for a false decision is thus 0.000001.” [3, p. 4]⁴
- *Database search case*: “There is no suspect. A search in the database is conducted, and exactly one person is found. That person is convicted on the basis of the same argument as that in scenario 1 [the probable cause case]. The conviction is of course, again, false, because the data of the true author are not stored in the database. What is the probability for such a false decision? The answer is somewhat more complicated than that in scenario 1 [the probable cause case]. An error occurs notably when exactly one person is found in the database. (...) We will find exactly one person with a probability of 0.368 (that is in approximately every third similar case), and this person will subsequently be convicted, even though the true author is not in the database. The probability for an error in scenario 2 [the database search case] is therefore considerably greater than in scenario 1 [the probable cause case].” [3, p. 4]¹

Based on this reasoning, Fimmers et al. argue [p. 4]:

“The simple evaluation using a likelihood ratio, as proposed by Taroni et al. [21] is appropriate for the first scenario [the probable cause case], yet produces an unjustifiably high number of false decisions in the second scenario [the database search case]”.¹

This is questionable, however, because a likelihood ratio in no way amounts to a categorical conclusion with respect to the process of individualization (i.e., the attribution of the trace to a single source to the exclusion of all other potential sources) [4]. In Fimmers et al.’s framework [3], every match results in a wrong individualization. Since every comparison of the crime stain’s profile with the profile of an individual in the population has a probability of 10^{-6} of leading to a match, every comparison has a probability of 10^{-6} of leading to a false individualization. It is therefore hardly surprising that Fimmers et al.’s probability of a

false individualization increases with the number of comparisons performed. In other words, one comparison in the probable cause case has a probability of

$$\gamma = 10^{-6}$$

of matching the crime stain’s profile, whereas one million comparisons in the database search case have a probability of

$$n\gamma(1 - \gamma)^{n-1} = 0.368$$

of leading to a match with the crime stain’s profile. This reasoning process consists of an unrealistic deduction based solely on the evidence (i.e., the observed match and the match probability of the crime stain’s profile). It is combined with an unusual definition of a population of potential sources, which does not contain the true source, and a definition of the decision as a categorical consequence of a match whenever $\gamma \leq 10^{-6}$.

There are many points to discuss regarding the arguments advanced by Fimmers et al. [3]. This paper treats the following three aspects:

- (A) The decision of ‘individualizing’ an individual as the source of a crime stain having a match probability of $\gamma = 10^{-6}$ in a population of 100 million potential sources ($N = 10^8$) after obtaining a single hit with this individual in a database containing 1 million of these potential sources ($n = 10^6$).
- (B) The assumption that the true source of the crime stain is *not* in the population considered by the investigators.
- (C) The conclusion that a probability of a false individualization is considerably greater in the database search case than in the probable cause case.

Throughout this paper, we will refer to these claims as points A, B and C.

2. Structure and contents of this paper

In this paper, we invoke decision theory to analyze the issue of how to decide to ‘convict’, or rather to ‘individualize’, the matching individual found in a database. The aim is to compare Fimmers et al.’s conclusions [3] (points A, B and C) with the results obtained from a decision-theoretic approach to the database search problem. Section 3 will present a decision-theoretic approach to the database search problem, using the visual representation of an influence diagram to clarify the relationships between the set of target decisions and the variables that pertain to the database search scenario formulated in its traditional version (which focused solely on probabilistic inference, rather than on decision making). This section uses the influence diagram to examine points A and B. We study the normative decision-theoretic framework in further detail in Section 4, in order to determine in which situations a rational decision maker would individualize the selected person. This section discusses point A in further detail. Section 5 examines the probability of a false individualization for a rational decision maker, and studies how this probability is regulated by the decision maker’s personal objectives and preferences. It concludes by comparing our results with the assertion made in point C. Section 6 summarizes the insights provided by this study, and discusses them with regard to Fimmers et al.’s argument [3].

Note that the decision of ‘convicting’ or ‘individualizing’ is made on the basis of the probability distribution over the propositions, and must therefore be on the same hierarchical level as the propositions in the hierarchy of propositions [16]. Fimmers et al.’s hypothetical case [3] addresses the decision at the crime level in

¹ Translation by the current authors, and words in square brackets added by the current authors.

² In a *probable cause case* (also referred to as the *confirmation case* [7,15]), a suspect found on the basis of information not related to the DNA profile has a profile that is subsequently found to match the crime mark’s profile.

³ In a *database search case* (also referred to as the *trawl case* [7,15]), a search in a database results in a single hit with an individual, who then becomes a suspect in this case because of the matching profile.

⁴ Translation by the current authors.

Download English Version:

<https://daneshyari.com/en/article/96491>

Download Persian Version:

<https://daneshyari.com/article/96491>

[Daneshyari.com](https://daneshyari.com)