

[翻訳]

## 「専門家の識別能力の階層的考察」

平 岡 義 博  
笹 倉 香 奈<sup>(訳)</sup>

Itiel E. Dror, A Hierarchy of  
Expert Performance, Journal of  
Applied Research in Memory and  
Cognition 5 (2016) 121–127.

### 要旨

専門家の識別能力は、専門家相互および専門家自身の信頼度とバイアス度を検証し、専門家が下す結論と観察を分けて検査することにより数値化することができる。これらのパラメータを用い「専門家識別能力の8段階 (HEP: A Hierarchy of Expert Performance)」を開発した。HEPを用いて刑事司法で重要な役割を担う高度な専門家領域である法科学の専門家の識別能力を評価し数値化する。HEPによって専門家のパフォーマンスを評価することで、その弱点を指摘すると同時に横断的に別の分野の専門家同士を比較することができる。またHEPは専門知識に対する理論的・実践的な洞察を可能にする。

日常業務の遂行が成功するかは人間のパフォーマンス能力にかかって  
る。航空、医療、法科学、金融、警察活動のように、専門家のパフォー

マンスが重要となる高度な専門領域であればなおさらである。このような領域やその他の領域においては、私たちの財産、健康、さらに生命もが専門家の手中に委ねられている。多くの専門家は長年の訓練を受け経験を持つ有能な人々であり、それゆえ需要が高く高収入でもある。

しかし能力のある専門家は、実際に良いパフォーマンスを発揮できているのか？専門知識や誰が専門家なのかさえ様々な見解や意見の相違もあって、定義づけることは非常に複雑かつ困難である (Feldon, 2007; Hoffman, 1996)。専門家能力の定量的評価に関する従来の研究は、特定の専門的分野に限られている。領域を横断して専門家パフォーマンスを評価する尺度や統一的な枠組みは存在しない。

領域横断的な専門家能力の重要な構成要素はバイアス度と信頼度にかかわる。バイアス度は無関係な文脈情報によって偏ることなく、関係ある情報に基づき判断する能力といえる。例えば「被疑者が犯行を自白した」「目撃者はその被疑者だと確認した」あるいは「捜査官は被疑者が犯人だと確信している」という情報を知った鑑定人は、科学的証拠（銃器、筆跡、音声、指紋など）が被疑者と一致すると思込み、誤って犯人を特定してしまうようなバイアスを持つ (Dror & Cole, 2010)。

バイアス度の問題に加えて基本的な信頼度の問題がある。無関係なバイアス文脈情報に曝されない場合でも、専門家が行う判断はどれだけ信頼度（一貫性や再現性）があるだろうか？別の鑑定人が同じ証拠物を検査して同じ結論にたどり着くのか（たとえば同じ被疑者を特定するか）？「バイアス度」は「信頼度」の上に無関係な文脈情報を付加して築かれるものであるので、信頼度の問題はバイアス度よりも基本的な問題なのである。「信頼度」は、バイアスの影響がない場合に専門家能力の一貫性について検査する。そのような意味で専門家能力にとってより基本的な要素といえる。

専門家相互のバイアス度と信頼度を比較することもできる。たとえば、(甲南法学'18) 58-3・4-114 (184)

二人の医用放射線技師が同じX線検査に基づいて同じ診断を下す程度を数値化することが可能である。日常的な意思決定において個人差は存在するものの、少なくとも専門家については、そのような相違（仮にそれが存在するとしても）が最小限であると我々は期待している。もし、二人の専門家の結論が異なり相反するならば、その時はどちらかが誤ったにちがいない。このようなことはより科学的で客観的な専門領域では重大な問題である。

専門家自身のバイアス度と信頼度を、より基本的なレベルで調べることがもできる。ここでは、異なる専門家間の能力を横断的に比較するのではなく、同じ専門家について異なる時期にバイアス度と信頼度をテストして比較する。つまり、専門家自身の信頼度は、同じデータを異なる時期に検査した場合、その専門家が同じ結論に到達するかどうかの程度を数値化する。専門家相互の信頼度の欠如は問題ではあるが、専門家自身における信頼度の欠如はより重大な問題である。専門家の意思決定はデータと専門知識から導かれるべきであるので、同じ状況を検査する専門家は一般的には同じ結論に達するべきである。

専門家自身のバイアス度は、同じ専門家が無関係な文脈情報の下で同じデータを検査した時、同じ結論に至る程度を数値化する。測定は専門家が無関係文脈情報によりバイアスを受けたかどうかである。同じケースが2回提示される真の信頼度とは異なり、ここでは無関係文脈情報が付加される。

専門家パフォーマンスについては、観察された情報とこれに基づき得られた結論とを区別する必要もある。たとえば医者は、顔色が悪く、意識障害があり、血圧が90/60であるという観察を行い、これらの観察から問題と治療方針を判断する。

この区別は重要であるが、これらの要素は相互に一体のものである。たとえばリスクをとるという判断は、第1にリスクの認知、第2に認知

リスクのレベルが判明した場合に次にとる行動から構成されている。もしこれら二つの要因が一体化されており分離されていないならば、リスクをとるという判断が行動を支えているようにみえたとしても、実際にはリスクをとるという判断ではなく、リスクの認識の欠如から行動が生じているかもしれないのである。

医療の領域では、「観察」と「結論」の違いはSBAR (Thomas, Bertram, & Jonson, 2009; Wacogne & Diwakar, 2010) の使用〔訳注：医療・介護現場におけるコミュニケーション技術〕により公式に分離されている。SBARとは Situation (状況)・Background (背景)・Assessment (判断)・Recommendation (提案)の頭文字であり、円滑なコミュニケーションのために使われる。「状況」と「背景」は「観察」を意味し、「判断」と「提案」は「結論」を意味する。

バイアス度と信頼度、専門家自身と専門家相互、観察と結論のような上述の要素と、信頼度はバイアス度よりも基本的であることと、結論が観察に基づくという階層的特徴を前提とすると、専門家能力は8段階に整理できる。以後、これを「専門家識別能力の階層 (Hierarchy of Expert Performance : HEP)」と名付け、重要な科学的評価が行われ高度に専門化された科学鑑定の実例の専門家を使っ、HEPの各レベルを概説する。

HEPの最上位(レベル8)は専門家相互の結論についてのバイアス度である。このレベルの専門家能力の検査は、異なる専門家が同じデータで意思決定する場合、無関係文脈情報によりバイアスを受けるかどうかを問うものである。例えばDNAの専門家は、事件の詳細な情報によってバイアスを受けることがあるか。これはDror and Hampikian (2011)により研究課題の一つにされた。あるDNA事案(実際の刑事事件)で、集団強姦の加害者の一人が答弁取引の一環として、他の被疑者に不利な証言をした。現場の証拠物は複数のDNAの専門家により検査されたが、この専門家たちは被疑者に不利な前述の証言を知っていた。さらにまた(甲南法学'18) 58-3-4-116 (186)

付加的で潜在的なバイアス情報は「DNA鑑定という補強証拠がなければ、司法取引に基づき得られた証言を法廷に提出することはできない」という事実であった。犯罪現場のDNA証拠物は混合物（複数の関係者のものを含む）であり、なんらかの判定や解釈が必要であった。DNA鑑定をおこなった二人の専門家はいずれも無関係文脈バイアス情報により「犯行現場のDNA混合物の一部がその被疑者に由来することは除外できない」と結論した。

無関係文脈情報が結論にバイアスを与えたかどうかをテストするため、Dror and Hampikian (2011) はそのDNA証拠を17名の専門家に見せた。ただし、先の2名の専門家が受けたような無関係文脈情報は提供しなかった。その結果、バイアス情報にさらされた専門家と同じ結論にいたったのは17名中1名のみであった。フィッシャーの正確確率検定に基づきp-値〔訳注：確率 (probability) p-値<有意水準 (例えば<0.05…過誤の確率が5%未満)〕と効果量〔訳注：どれくらいのデータでどれくらいの効果があるか〕の計算は、p-値=0.018(片側検定)、効果量=0.49であり、よってDror and Hampikian (2011) は「事件の無関係な文脈情報がDNA証拠の解釈に影響を及ぼした可能性がある」と結論づけた。

Murrie, Boccaccini, Guarnera, and Rufino (2013) は、同じ証拠物を異なる専門家に示すことにより無関係文脈情報を操作したが、これは誰のために働くかという専門家の意識を操作するものであった（ある専門家には「弁護側に依頼された」と告げ、他の専門家には「検察側から依頼された」と伝えた）。この無関係文脈情報は提示された4事件のうち3件で専門家の結論にバイアスを与えた（3事件に対するコーエンのd値〔訳注：平均値の差／標準偏差。効果の大きさを表す。d値が小さい場合は差を有意にするためにはより多くのデータが必要となる〕は0.55～0.85で有意な結論）。このように無関係文脈情報がバイアス効果をもたらすことがまたもや示された（何が無関係文脈情報であるかの議論は

National Commission on Forensic Science, 2015 参照)。

HEP のレベル 8 は専門家相互のバイアス効果に関するものである。これに対してレベル 7 では専門家自身のバイアスを検査する。レベル 7 がレベル 8 の下位にあるのはより基本的な問題を検査するためである。つまり、同じ事案が異なる無関係文脈バイアスの下で提示された場合、同じ専門家が同じ（または異なる）結論に到達するかという問題である。この問題を調査するにあたり、指紋専門家に過去に自らが鑑定した事案が提示された。彼らはこれらの事件について再度鑑定するよう指示されたが、自らが過去に扱った指紋であることは知らされなかった。ただし、無関係文脈情報が仕掛けられた。例えば、もし専門家が過去に「指紋が合致した」と結論していたなら、今回は「別人が自白した」とか「被疑者には確かなアリバイがある」などの文脈の下で提示された。このような操作は二つの研究 (Dror & Charlton, 2006; Dror, Charlton, & Peron, 2006) で使用され、そのメタ分析〔訳注：統計解析された結果を収集し、様々な角度から比較統合し、統計的有意性のある差異の検出が可能となる（研究の抽出、データ統合、効果量の算出）〕が Dror and Rosenthal (2008) で報告された。

指紋専門家は無関係文脈情報によりバイアスを受ける可能性があることがわかった。無関係文脈情報の下に同じ試料が再度提示された場合、専門家は17%から80%の場合に結論を変えたことが示された（この所見と提案された解決法は Dror & Cole, 2010; Kassin, Dror, & Kukucka, 2013 参照）。専門家の結論におけるバイアス効果の存在と強度は次の3つの要因に依存していた。

## 1 バイアス情報の強度

一部の無関係文脈情報は他の情報よりもバイアスを生じやすい。例えば確固たるアリバイは「刑事はこの被疑者が罪を犯したとは確信している（甲南法学'18）58-3・4-118（188）

ない」という情報より強い影響を及ぼす。多くの法科学鑑定者は、自分たちは客観的であり、バイアスのある文脈情報から影響は受けないと考えている（例えば「捜査資料を読みたいという鑑定人や鑑定前や鑑定中に捜査員と協議したいという鑑定人。そのような興味は判断を変えることもなく仕事を楽しむために許されるある種の自己満足をもたらすにすぎない。確かに法科学者によって行われる鑑定業務は退屈で日常的であり、多くの者にとっては面白くもバラエティ豊かなものでもなく、興味のない仕事かもしれない」（Butt, 2013, p.60. 付点は引用者）。最近、無関係文脈情報が鑑定結果にバイアスを与えうるという認識が高まっており、法科学専門家からそのような無関係文脈情報を遠ざけることを義務付ける動きがある（Science (Servick, 2015), Nature (Spinney, 2010), 米国法科学委員会の文書「法科学分析がタスク関連情報に基づくことの保証」2015, 英国法科学監視指導文書「法科学鑑定に関する認知バイアス効果」, 2015）。

## 2 決断の困難性

結論が容易で明確である場合、バイアス情報は最終結論に影響を与えないが、結論が微妙な場合には影響を与える。犯行現場からの証拠が量・質ともに少ない場合は専門家の結論に影響する無関係文脈情報が入りこみやすい余地がある。バイアス情報の強度は、決断の困難性ととも、「バイアス危険領域（バイアスが法科学の結論に最も影響を与えそうな事例）」を決定する。

## 3 バイアスの方向

法科学専門家を決定的な「合致」結論（これは公判で説明することを求められる）ではなく、より曖昧な「鑑定不能」の結論に結び付くようバイアスをかけることは容易である。にもかかわらず、同じ指紋が無関

係文脈情報の下で同じ専門家に提示された時、当初の確信に満ちた「合致」結論が「鑑定不能」や「不合致」になってしまう、というデータも示されている。

HEP のレベル 6 では、バイアス度ではなく、より基本的な信頼度に関するパフォーマンスを測定し数値化する。バイアスとなる無関係文脈情報がなくても、専門家たちが出す結論は相互に一致するだろうかという問題である。DNA 鑑定の専門家について実施された多くの研究はこの問題について検討を行い、これらの検討によって DNA 鑑定専門家の信頼度に大きな問題があることが明らかになった（複雑な DNA 鑑定が必要な場合、例えば複数人から由来する DNA 混合物が扱われる場合、多くの判断が DNA 鑑定専門家の裁量により行われている（Starr, 2016））。この問題を最初に指摘した研究（Dror and Hampikian, 2011）では、17 名の専門家に同じ混合 DNA 試料を提示した。17 名の専門家は同じ研究所に所属し、同じ手続き・プロトコルで鑑定が行われたが、その結論は異なるものになった。すなわち、12 名の専門家は、DNA 試料が被疑者に由来しないと結論し、4 名は判定不能、1 名は被疑者に由来する可能性があると結論した。

米国標準技術局（National Institute of Standards and Technology）により（裁判所でもしばしば用いられる）確立された統計処理ツールに基づきこの DNA 鑑定についての所見が再確認された（Coble, 2015）。Dror and Hampikian（2011）と同じように Coble（2015）でも、同じ法科学研究所の中においてさえ、同じ統計手法を用いた法科学の結論に明らかな相違と変動がみられたのである（Starr, 2016）。同様の結果は足こん跡鑑定のような他の専門家領域でも見られた（Majamaa & Ytti, 1996）。

HEP のレベル 5 では、専門家自身の結論の信頼度を検査する。専門家相互の相違（レベル 6）の測定よりもっと基本的なレベルの相違、すなわち専門家自身の信頼度を数値化するものである。Dror and Charlton（甲南法学'18）58-3・4-120（190）

(2006; さらに Dror & Rosenthal, 2008) は指紋鑑定について、(文脈やバイアス度を操作しない場合でさえ) 同じ資料が提示された時、同じ指紋専門家が異なる結論に至る確率が8%あったという結論を得た。

連邦捜査局 (Federal Bureau of Investigation, FBI) の研究は、Dror とその共同研究者による最初の研究結果を再確認した (Ulery, Hicklin, Buscaglia, & Roberts, 2012)。FBI の研究では、72名の指紋鑑定者に25対の指紋を約7ヶ月の期間をおいて2回検査させた。指紋鑑定者は同じペアの指紋の結論を10%の確率で違ったという調査結果であった。

これまで述べた HEP のレベル5～8は、専門家の「結論」に関するものである。結論はデータと証拠の観察のみに基づいて下される。例えば指紋専門家は、犯人の指紋など犯罪現場で採取される潜在指紋〔訳注: 肉眼で見えない指紋〕の特徴点 (指紋の特徴—図2参照) を被疑者指紋の特徴点と比較し照合する。現場の指紋と被疑者の指紋とが同じ起源のものといえるかどうかについて、これらの指紋が十分似ているかを比較し判断する。しかし、その結論は指紋の特徴点の観察がどのように行われたかということに依存する。

HEP のレベル5～8 (「結論」) とレベル1～4 (「観察」) の違いは、ある意味で記述統計学と推測統計学に類似している。また、医療分野で使用される SBAR (前述の状況・背景・評価・提案) にも似る。レベル1～4はデータ観察における専門家能力に関するものであり、専門家相互および専門家自身のバイアス度と信頼度を測定することで測定される。

図1のように、HEPは観察レベルを結論レベルの下位に配置する。観察は結論よりも基本的だからである。観察は専門家が到達すべき結論の土台となるものである。結論が観察を導き出す場合には (その逆よりも) 確証バイアスが入りこみやすい。それ故、概念的にも実質的にも観察が結論よりも先に行われるべきである。さらに、犯罪現場の証拠物から観察を始め、その後被疑者からの資料を観察しなければならない。

この順番に従わない場合には、逆行的な順番で行われることになり危険である。すなわち（無関係文脈情報からもたらされる）示唆された結論から始め、次に証拠物の中に被疑者の特徴を見出すような判断が行われるのである（つまり、証拠物から被疑者に至るのではなく、被疑者から証拠に至ってしまう）。

このようなバイアスに対抗することは比較的容易である。つまり（無関係文脈情報にさらされずに）観察から始め、証拠から被疑者へと進んでいくことである。線形逐次顕在化（LSU：Linear Sequential Unmasking）手法はまさにその通りのことを実行するものである（Dror et al., 2015）。LSU手法は、犯罪現場の証拠物と被疑者とを切り離して、初めに犯罪現場の証拠物を観察することを義務づけている。法科学専門家は

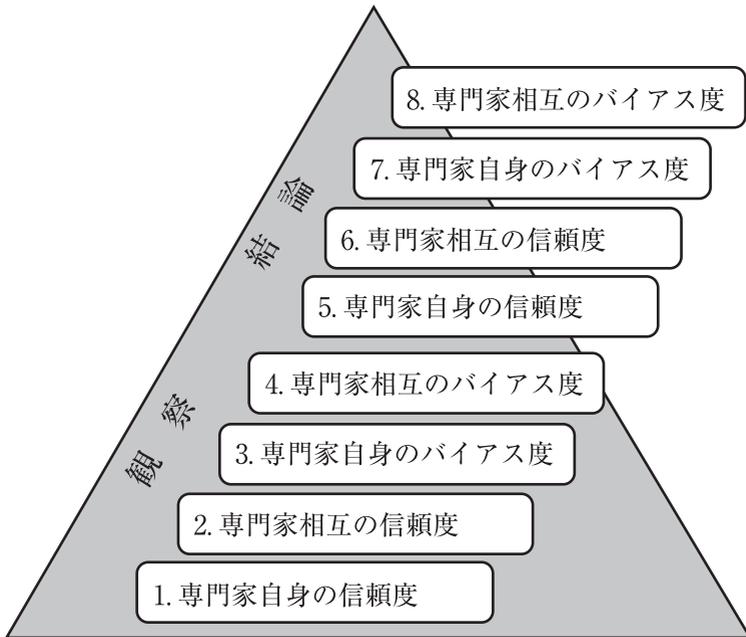


図1 専門家識別能力の階層（HEP）

始めに実際の証拠物を検査し、観察した事項を記録する必要がある（例えば DNA 鑑定では現場遺留物からの DNA のプロファイル〔訳注:DNA 型を記録したエレクトロフェログラム〕を検査しピークがどこにあるかを特定する）。その後でターゲットとなる被疑者の資料を観察し検査する（DNA の場合には、被疑者の DNA プロファイルを検査する）。そして正しい順番（最初に犯罪現場、その後に被疑者）で両者の DNA プロファイルを観察した後、両者を比較して結論に至る（Dror et al., 2015）。しかし専門家はどの段階でも無関係文脈情報にさらされるべきではない。なぜなら LSU は当初の不適切な分析を補正し（Ulery, Hicklin, Roberts, & Buscaglia, 2014）、より効果的な視覚的探索を可能にするためにターゲットとなる被疑者を明らかにした後で（制限付きではあるが）一定の修正を行うことを認めているからである。

観察の数値化の最初に出てくるレベル 4 は、データの観測において無関係文脈情報によりバイアスの影響を受けるかを検査するものである。一つのバイアス情報源は事件情報そのものである（例えばレベル 7、8 で示したような他の情報筋からの被疑者に不利な証拠の存在）。しかしながらバイアス情報には 5 つの異なる情報源が存在する（Dror, 2015; Dror et al., 2015）。前述したようにもう一つのバイアス情報は「ターゲット」〔訳注:現場遺留資料（現場に残された指紋や血痕など）の中から被疑者に関係するものを探したり、検索データベースから合致データを探し出したりする時、そのターゲットは被疑者に照準が当てられる〕、つまり、当該証拠物が合致されるべき対照資料である（被疑者の指紋や DNA プロファイルなど;Thompson, 2009）。ここで現場に遺留された証拠物は「ターゲットである被疑者」という文脈の下で観察される。このことによって、実際の証拠に対する観察にバイアスがかかる。Dror らの研究（2011）はそのような影響が本当にあるかを調べた。指紋鑑定の専門家に対し、犯罪現場の潜在指紋のすべての特徴点（図 2）を計数さ

せた。第一に現場指紋を単独で、第二に被疑者がターゲットであるというバイアス文脈情報の中で、である。結果は、犯罪現場遺留の潜在指紋が「ターゲットである被疑者」という文脈で提示された時、専門家は異なる数の特徴点を観測するというバイアス効果が示された。

バイアスにより特徴点の数が異なる観察が行われるというこの結果は、今後、被疑者指紋の比較が行われる場合に示唆を与えてくれるだけではない。そもそもある証拠が元々比較対照として十分なのか、という判断に対しても示唆を与えてくれる。つまり、現場遺留の潜在指紋（数が少なく質の悪い情報しか保有しないことも多い）がそもそも照合に値するかどうかということである。しばしば潜在指紋は、被疑者と比較できるほど十分な情報を保有しないのである。Fraser-Mackenzie, Dror, and Wertheim (2013) は、潜在指紋の適切な鑑定ができるかという判断に、バイアスが影響するかどうかを調査した。この研究で、専門家は潜在指紋を単独で提示するという条件と、ターゲットである被疑者指紋との照合と同時に提示される条件のいずれかのもとで潜在指紋を観察した。専門家がバイアスをもって異なる数の特徴点を観測したという研究と同じ

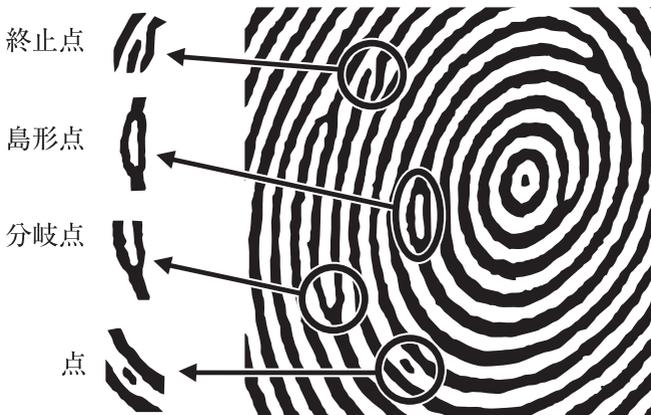


図2 指紋の皮膚紋理に現れる特徴点（紋理が分離または途切れる場合）

く、この研究は潜在指紋が照合に適切かどうかという観察が、バイアスを受けることを明らかにした。

レベル4は異なる専門家相互の観察におけるバイアス度を検査したが、HEPのレベル3ではより基本的な専門家自身の観察のバイアス度を検査する。つまり、無関係文脈情報の下に証拠物が提示された時、同じ専門家がその証拠物について異なった観察をするかという問題である。

Earwaker, Morgan, Harris, and Hall (2015) による研究は、専門家自身の中でもバイアスを受けた観察を行うことがあるという証拠を見出した。同一の潜在指紋が照合に十分な情報を有しているかどうかという判断は、無関係文脈情報によってバイアスを受けていたのである。照合に不十分で適当でない潜在指紋が、重大事件（殺人事件など）の文脈情報の下に提示された時には十分なものとして観察されることが多く、一方で照合に十分で適当な潜在指紋が重要度の低い事件（窃盗事件など）の文脈情報の下に提示された時には、不十分なものとして観察されることが多くなった。このバイアス情報のために、十分なものであるとして観察された潜在指紋の34%は実際には照合が十分にできる質のものではなく、照合に不十分で不適当と観察された潜在指紋の33%が実際は照合が十分にできる良質なものであった (Earwaker et al., 2015)。

潜在指紋に関する観察がバイアスを受ける危険性がわかる良い例は、R. v. Smith 事件（2011）であろう。この事例において、イングランド・ウェールズ控訴院は殺人事件の有罪判決を破棄した。殺人事件の「ターゲット」たる被疑者の存在に潜在指紋の観察が影響され、そのことによって指紋証拠がバイアスを受けたものになっていたという理由であった。現場遺留の潜在指紋が被疑者と切り離されて（ターゲットたる被疑者の指紋の前に）観察された際には、指紋専門家は「有意な照合を行うには不十分な特徴点しかなかった」としていた。しかし「被告人が起訴された後には……鑑定人は指紋の稜線の流れと12の稜線の特徴は被告人の人

差指の指紋に合致したと結論した」。つまり、鑑定者は潜在指紋についての彼自身の観察を「照合に不十分で不適當」から「適當」に「変え」たのである（そして、被疑者の指紋と合致すると結論づけたのである）

（R. v. Smith, 2011, 第14段落と第15段落）。この問題は、証拠から被疑者を割り出すのではなく、被疑者から証拠に向かうという逆行的な順番での手続を採用してしまったことから生じる（この点は Linear Sequential Unmasking (LSU) 手法のところで前述した）。

HEPの下位に行くと、最も基本的な専門家パフォーマンスの測定が行われることになる。レベル2はデータの観察（バイアスを与える文脈情報が無い状態）の専門家相互の基本的な信頼度に関連がある。バイアスを与える文脈情報がない指紋照合において（レベル3と4はバイアス文脈の下で行った）、異なる指紋専門家は、同じ数の特徴点を検出するものなのか、という問題である。Drorら（2011）はこの問題を調査した。表1は指紋専門家相互で信頼度と整合性が欠如していることを示す。

表1のデータが示すように、バイアス文脈情報が提示されない場合でも、専門家たちの指紋照合結果は相互に異なっていることが認められる。特徴点の計測数が大きくばらつく（専門家が計測した特徴点の数が広い範囲にばらつき、標準偏差が大きくなる）だけでなく、照合を行うための臨界閾値〔訳注：判別基準の上限值〕が観察ごとに変わっている。照合に必要な特徴点の定数〔訳注：合致基準値〕は多くの国では決められていないが、7点か7点未満の特徴点では、合致すると結論づけるには十分な数とはいえない。そのような意味で、表1の結果は憂慮すべきものがある。指紋資料のD, E, F, G, I, J（指紋の半数以上）については7点かそれ以下の特徴点を観測した専門家がいる一方で、他の専門家は同じ指紋資料について明らかに多くの特徴点（少なくとも12点）を観測している。このように、合致するという結論はどの専門家が犯罪現場からの潜在指紋を観察するかによって決まってしまうのである。

表1 専門家相互の指紋観察

専門家	潜在指紋									
	A	B	C	D	E	F	G	H	I	J
Expert 1	22	9	15	8	9	3	8	11	7	10
Expert 2	21	11	25	7	10	9	9	10	6	5
Expert 3	19	9	18	10	7	9	15	19	6	6
Expert 4	21	21	29	14	12	9	8	9	4	8
Expert 5	17	16	15	11	16	9	7	12	5	5
Expert 6	20	14	22	9	10	7	13	18	7	9
Expert 7	22	17	15	10	10	8	11	24	8	11
Expert 8	9	9	19	6	9	8	18	16	9	10
Expert 9	30	15	25	10	12	12	19	22	12	17
Expert 10	25	13	18	13	12	10	13	15	7	10
Min	9	9	15	6	7	3	7	9	4	5
Max	30	21	29	14	16	12	19	24	12	17
SD	5.49	4.01	4.93	2.49	2.45	2.32	4.25	5.15	2.23	3.54
Range	21	12	14	8	9	9	12	15	8	12

註 同じ指紋（A～J）を観察した異なる指紋鑑定専門家の観測した特徴点数は、一貫性を欠く

例えば、専門家2が指紋資料Jを検査すれば、わずか5個の特徴点のため合致したとの結論に至らないが、専門家9が同じ指紋をみれば17の特徴点により合致照合がありうるという具合である。専門家の特徴点観察の変動はSwofford, Steffan, Warner, Bridge and Salyards (2013) もまた数値化し指摘している。

HEPの最下位のレベル1は最も基礎的な専門家パフォーマンスを検査するものである。同じ専門家が同一の証拠物を見て同じデータを観測するであろうか、という問題である。そのような専門家自身による測定は専門技能の基本中の基本である。Drorら(2011)はこの問題を研究し、指紋鑑定者たちの観察は相互に一致しない(レベル2)だけでなく、同じ専門家自身でも常に自分自身の観察が一致するとも限らないことを明らかにした(表2参照)。

すなわちバイアス文脈情報がなくても、同じ専門家は同一の指紋資料から異なるデータを観測する。同じ専門家が1回目と2回目に観測した特徴点の数の差は9個までになる、というデータが示された。専門家自

表2 専門家自身の指紋観察

専門家	潜在指紋									
	A	B	C	D	E	F	G	H	I	J
Expert 1	1	1	4	1	1	2	3	2	0	1
Expert 2	8	3	5	1	1	2	2	5	2	2
Expert 3	1	3	3	3	6	4	9	9	1	2
Expert 4	2	3	2	5	0	1	1	0	0	1
Expert 5	6	2	2	3	4	1	3	3	0	3
Expert 6	9	4	2	1	4	6	0	5	1	1
Expert 7	0	4	5	2	4	3	3	7	0	0
Expert 8	3	1	4	0	6	2	1	4	2	0
Expert 9	4	3	9	0	4	4	3	1	1	3
Expert 10	1	0	0	1	4	1	4	1	0	0
MEAN	3.5	2.4	3.6	1.7	3.4	2.6	2.9	3.7	0.7	1.3

註 各専門家が同一の指紋を異なる機会に2回検査した時の特徴点計測数の差異（0は一貫性があることを示す）

身の能力の信頼度（1回目と2回目の差が0となる一貫性）は全体の16%とわずかであった。合致判断の閾値を超える場合には、この信頼性の欠如という問題は重大である。そしてデータによれば、どの専門家についても閾値を超える判断が行われる場合が実際にある（例えば、専門家2が指紋Hについて1回目に特徴点を8点観測したが2回目は13点観測、専門家3が指紋Gを8点から17点へ、……専門家7が指紋Bを9点から13点へ—Dror et al., 2011 参照）。よって合致の可能性は、どの専門家が潜在指紋を観察するのか（HEPのレベル2）に依存するだけではない。同じ専門家しか関与しない時でさえ、ある時は「合致」のために十分な特徴点を観測する一方で、「合致」にできない別の時にははるかに少ない数の特徴点を観察するのである。特徴点観測における専門家自身の中の変動は、Swofford, Steffan, Warner, Bridge and Salyards(2013)でも数値化され指摘されている。

つまり専門家の識別能力の階層（HEP）は、データの観察、そして結論について、専門家のパフォーマンスの測定と数値化の枠組みを提供する。無関係文脈情報によるバイアスの影響という高次のレベルだけでなく（甲南法学'18）58-3・4-128（198）

く基本的なレベルである専門家の信頼度について、HEPによりその程度を数値化できる。HEPはまた、専門家のパフォーマンスが専門家相互で変動する場合と、専門家自身で変動するようなより基本的なケースを区別する。この枠組みは明瞭な8段階に定義され、それぞれについて専門家のパフォーマンスの異なる側面を数値化できる。

HEPの枠組みは、法科学の専門家の能力を分析し、これを測定するのに使用できる。全米科学アカデミー（2009）による法科学の領域への批判以来、法科学の専門家のパフォーマンスについて多くの研究がみられるようになった（Found, 2015）。HEPを用いることにより、これまでの研究を整理・概念化し、明確な理論的枠組みを提供することができる。また、専門家のパフォーマンスの弱点を明らかにし、それらの弱点を解消するための手続きを明確にするという応用も可能になるであろう。

この論文ではHEPを法科学の専門家領域に用いたが、他の専門家領域でも応用可能である。HEPはまた、専門家領域を横断的に8レベルの階層で比較することにより、未熟者の評価だけでなく専門知識の評価にも応用できる。階層は、専門家領域を超えて信頼度とバイアス度問題に焦点をあてるものであり、有効性〔訳注:「有効性」は測定値が真の値や真実に近い程度、これに対し「信頼度」はデータの繰り返し精度として区別〕の問題は取り扱わなかった。妥当性の問題は、ある専門家領域内の問題だからである。HEPは専門知識に対する実践的で理論的な洞察を提供するものである。

利益相反について

著者には利益相反がない。

\*原著翻訳は、Elsevier and Copyright Clearance Centerのライセンス(No.4175240190727)に基づき行った。

References

- Butt, L. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions — Commentary by a forensic examiner. *Journal of Applied Research in Memory and Cognition*, 2 (1), 44–45.
- Coble, M. (2015). *Interpretation errors detected in a NIST inter-laboratory study on DNA mixture interpretation in the US (MIX13)*. Washington, DC: Presentation at the international symposium on forensic science error management – detection, measurement and mitigation.
- Dror, I. E. (2015). Cognitive neuroscience in forensic science: Understanding and utilizing the human element. *Philosophical Transaction of the Royal Society B*, 370, 20140255.
- 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, 10–17.
- Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, 56 (4), 600–616.
- Dror, I. E., & Charleton, D., & Peron, A., (2006). Contextual information renders experts vulnerable to making erroneous identifications, *Forensic Science International*, 156 (1), 74–78.
- Dror, I. E., & Cole, S. (2010). The vision in blind justice: Expert perception, judgment and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review*, 17 (2), 161–167.
- Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, 51 (4), 204–208.
- Dror, I. E., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences*, 53 (4), 900–903.
- Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Context Management Toolbox: A Linear Sequential Unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, 60 (4), 1111–1112.
- Earwaker, H., Morgan, R. M., Harris, A. J., & Hall, L. A. (2015). Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need? *Science & Justice*, 55 (4), 239–247.
- Feldon, D. F. (2007). The implications of research on expertise for curriculum

- and pedagogy. *Educational Psychology Review*, 19 (2), 91–110.
- Forensic Regulator. (2015). *Cognitive bias effects relevant to forensic science examinations*. Available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/470549/FSR-G-217\\_Cognitive\\_bias\\_appendix.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470549/FSR-G-217_Cognitive_bias_appendix.pdf)
- Found, B. (2015). Deciphering the human condition: the rise of cognitive forensics. *Australian Journal of Forensic Sciences*, 47 (4), 386–401.
- Fraser-Mackenzie, P., Dror, I. E., & Wertheim, K. (2013). Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments. *Science & Justice*, 53 (2), 144–153.
- Hoffman, R. R. (1996). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds.), *Exploring expertise* (pp. 81–100). Edinburgh, Scotland: University of Edinburgh Press.
- 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.
- Majamaa, H., & Ytti, A. (1996). A survey of the conclusions drawn of similar footwear cases in various crime laboratories. *Forensic Science International*, 82, 109–120.
- Murrie, D., Boccaccini, M., Guarnera, L., & Rufino, K. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, 24, 1889–1897.
- National Academy of Sciences. (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC: National Academies Press.
- National Commission on Forensic Science. (2015). *Ensuring that forensic analysis is based upon task-relevant information*. Available at: <https://www.justice.gov/ncfs/file/818196/download>
- Smith, R. v. (2011). *EWCA (Crim) 1296 (Eng.). Case No:2009/03393/C1. In The High Court Of Justice, Court Of Appeal of England and Wales (Criminal Division)*.
- Starr, D. (2016). When DNA is lying. *Science*, 351, 1133–1136.
- Servick, K. (2015). Forensic labs explore blind testing to prevent errors. *Science*, 349 (6247), 462–463.
- Spinney, L. (2010). Science in court: The fine print. *Nature*, 464, 344–346.
- Swofford, H., Steffan, S., Warner, G., Bridge, C., & Salyards, J. (2013). Inter- and

- intra-examiner variation in the detection of friction ridge skin minutiae. *Journal of Forensic Identification*, 63 (5), 553–570.
- Thomas, C. M., Bertram, E., & Johnson, D. (2009). The SBAR communication technique. *Nurse Educator*, 34 (4), 176–180.
- Thompson, W. C. (2009). Painting the target around the matching profile: The Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk*, 8, 257–276.
- Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, 7 (3), e32800.
- Ulery, B. T., Hicklin, R. A., Roberts, M. A., & Buscaglia, J. (2014). Changes in latent fingerprint examiners markup between analysis and comparison. *Forensic Science International*, 247, 54–61.
- Wacogne, I., & Diwakar, V. (2010). Handover and note-keeping: the SBAR approach. *Clinical Risk*, 16 (5), 173–175.