

De wankle kritiek op het forensisch handschriftonderzoek

Reinoud Stoel¹, Charles Berger¹, Elisa van den Heuvel¹, Wil Fagel¹

Professor Merckelbach heeft het in zijn artikel “De wankle basis van de forensische handschriftkunde”² uitgebreid over een aantal zaken dat die wankelheid aan zou moeten tonen. De voorbeelden komen uit verre landen en verre verledens, en zijn ook daardoor niet altijd even relevant voor de huidige praktijk bij het Nederlands Forensisch Instituut (NFI). In dit artikel willen we echter ingaan op de problemen die er *wel* zijn, en de oplossingen daarvoor.

Inleiding

De zaken die Merckelbach aanhaalt uit de VS en Frankrijk van een eeuw terug kenden vele problemen. De destijds veronderstelde pijlers onder het handschriftonderzoek waar Merckelbach op focust vormden slechts één van die problemen. Dat elk handschrift uniek is en dat een handschriftonderzoeker in staat zou zijn die uniciteit te herkennen is *niet* de basis van de huidige praktijk bij het NFI. Dit overigens niet omdat niet aan te tonen zou zijn dat ieder handschrift uniek is. Sterker nog: geen twee handschriften kunnen identiek zijn, omdat het dan niet twee maar één handschrift zou betreffen.³ Het probleem is daarom juist dat ook meerdere handschriften van dezelfde schrijver nooit identiek zullen zijn. Op wat dan wel de basis voor het handschriftonderzoek vormt zullen we verderop in dit stuk terugkomen.

De gegeven voorbeelden van missers hangen vaker met andere (ethische) problemen samen dan met een ontbrekende wetenschappelijke basis. De voorbeelden

-
1. De auteurs werken allen bij het Nederlands Forensisch Instituut (NFI).
 2. Merckelbach, H. (2010). De wankle basis van de forensische handschriftkunde. *Nederlands Juristenblad*, 2010, 7, 413-416.
 3. Ludwig Wittgenstein zei, ‘Beiläufig gesprochen: Von *zwei* Dingen zu sagen, sie seien identisch, ist ein Unsinn, und von *Einem* zu sagen, es sei identisch mit sich selbst, sagt gar nichts.’, in *Logisch-Philosophische Abhandlung*, Wilhelm Ostwald (ed.), Annalen der Naturphilosophie, 1921 (14), 5.5303.

die Merckelbach geeft zijn daarnaast soms verkeerd begrepen of niet goed doorgrond. Zo bleken de handschriften in de ‘Hitlerdagboeken’ wel degelijk van dezelfde hand als het referentiemateriaal. De misser in die zaak was juist dat de vervalser ook (een deel van) het referentiemateriaal geproduceerd had. De desbetreffende deskundigen hadden dus in die zin gelijk wanneer zij stelden dat de geschriften (dagboek en referentie) een gemeenschappelijke auteur hadden. Maar het zou te gemakkelijk en onterecht zijn om op grond daarvan kritiek op het forensisch handschriftonderzoek weg te wuiven.

Het artikel van Merckelbach eindigt vrij abrupt met de (helaas) niet nader uitgewerkte suggestie om een evidence line-up uit te voeren, dit als kennelijke oplossing voor de aangevoerde problemen. We zullen in dit stuk de door Merckelbach besproken issues met betrekking tot forensisch handschriftonderzoek op een rijtje zetten, en laten zien dat er oplossingen voor bestaan. Daarbij zal blijken dat een aantal van die issues en oplossingen daarvoor door Merckelbach verward worden. Ook zal duidelijk worden dat de door hem voorgestelde oplossing van de evidence line-up bepaald niet zo na te streven is als door hem, en soms door anderen, wordt gepretendeerd.

Basisprincipes en logica

We meldden al eerder dat de genoemde pijlers onder het handschriftonderzoek (dat elk handschrift uniek is en dat een handschriftonderzoeker in staat zou zijn die uniciteit te herkennen, ook wel ‘discernable uniqueness’ genoemd) niet de basis van de huidige praktijk op het NFI vormen. De handschriftdeskundige levert ook geen oordeel over het al dan niet van dezelfde schrijver afkomstig zijn van twee handschriften. Dat kan hij ook helemaal niet omdat de kans dat de handschriften dezelfde bron hebben ook van andere informatie afhangt dan alleen die handschriften zelf. De handschriftdeskundige onderzoekt en vergelijkt een betwist en een referentiehandschrift afkomstig van de verdachte, en maakt een inschatting van de waarschijnlijkheden van de door hem waargenomen overeenkomsten en verschillen onder de hypothese van gelijke en de hypothese van verschillende herkomst. Over de principes die ten grondslag liggen aan deze methode, die ook wel bekend staat als de ‘Likelihood Ratio’ methode, is eerder in dit blad geschreven.⁴

4. Berger, C.E.H. (2010). Criminalistiek is terugredeneren. *Nederlands Juristenblad*, 2010 (13), 784-789.

De kern van de Likelihood Ratio methode is dat de (handschrift)deskundige zich beperkt tot een uitspraak over de bewijskracht van zijn waarnemingen in het licht van de opgestelde hypothesen. Er wordt, extreme gevallen daargelaten, geen uitspraak gedaan over de waarschijnlijkheid van de hypothesen. Dit laatste is namelijk een taak van de rechter. In een zaak waar het gaat om een bedreiging, bijvoorbeeld, kan de handschriftdeskundige dus niets zeggen over de waarschijnlijkheid dat de verdachte de dreigbrief heeft geschreven. Hij zou hiervoor veel meer informatie nodig hebben dan tot zijn expertise behoort. Immers, ook tactische informatie en ander bewijs draagt bij aan deze waarschijnlijkheid. De handschriftdeskundige kan en zal zich alleen uitlaten over (1) hoe goed de kenmerken van het handschrift in de dreigbrief passen binnen de natuurlijke variatie van het handschrift van de verdachte, en (2) over hoe zeldzaam deze kenmerken zijn in de relevante populatie van mogelijke daders.⁵ Met andere woorden: hoe waarschijnlijk is deze waarneming wanneer het door dezelfde schrijver is geschreven, en hoe waarschijnlijk is de waarneming wanneer een willekeurig ander het heeft geschreven. Het quotiënt van deze twee waarschijnlijkheden definieert de bewijskracht, of kortweg, de LR (Likelihood Ratio).

Objectivering

Hoewel we in dit stuk voornamelijk zullen spreken over de ‘handschriftdeskundige’ en ‘het handschriftonderzoek’ gelden de genoemde issues en oplossingen veel breder. Een breed levend streven in de forensische wetenschap is het objectiveren van analyse, vergelijking en interpretatie door verkleining van de subjectieve component daarin. Dit is een grote uitdaging voor de forensische wetenschap waarbij ten eerste de te onderzoeken kenmerken exact worden gedefinieerd, en objectief meetbaar worden gemaakt. Vervolgens moet een vergelijkingsmethode gedefinieerd worden die de mate van overeenkomst (of verschil) tussen twee sets kenmerken oplevert. Voor het objectiveren van de interpretatie moet vervolgens worden bepaald welke mate van overeenkomst (of verschil) wordt verwacht bij materiaal uit dezelfde en bij materiaal uit verschillende bron. Aan deze uitdaging wordt volop gewerkt, maar voor de ene forensische discipline zijn veel eerder resultaten te verwachten dan voor de andere. Het

5. Wanneer het mogelijk om niet spontaan handschrift gaat moet uiteraard ook rekening worden gehouden met nabootsing of juist verdraaiing van handschrift.

forensische handschriftonderzoek is een van de lastigst te objectiveren soorten onderzoek.

Ringonderzoeken

Subjectief staat echter niet gelijk aan onbetrouwbaar. Een subjectief oordeel kan heel betrouwbaar zijn. Onder subjectief verstaan we hier conclusies die niet (alleen) gebaseerd zijn op harde data, maar (ook) op een persoonlijke inschatting door de deskundige op basis van zijn kennis en ervaring.

Regelmatig wordt onderzoek gedaan naar subjectieve oordelen in zogenoemde ringonderzoeken. Bij dit type onderzoek wordt de deskundige getest met vele sets handschriften die telkens al dan niet van dezelfde schrijver stammen. Anders dan in het echte zaakonderzoek wordt de deskundige nu gedwongen te kiezen uit ‘zelfde schrijver’, ‘verschillende schrijver’ en ‘geen oordeel’ om zo zijn resultaten direct te kunnen vergelijken met de bekende juiste antwoorden. In tegenstelling tot bij zaken, is de werkelijke uitkomst bij ringonderzoeken natuurlijk wel bekend, en zal men trachten zo moeilijk mogelijke zaken in het onderzoek te betrekken.

Het doel van een ringonderzoek is om instituten en onderzoekers met elkaar te vergelijken. Men kan zodoende al dan niet laten zien dat men competent is, en onderling consistent is in het oordeel. Het doel van een ringonderzoek is niet om een foutenmarge toe te kennen aan categorische oordelen die niet gegeven worden (zoals ‘deze handtekening is een vervalsing’). In de literatuur wordt vaak een foutpercentage van rond de 7% genoemd, en zo ook door professor Merckelbach. Een dergelijk percentage zegt dus heel weinig over het foutenpercentage in de praktijk, waarin men tot een ander soort conclusies komt en waarin het onderzoek altijd door een tweede deskundige geschaduwd wordt. Het zegt al helemaal niets over wetenschappelijke basis.

Ongewenste invloeden

Cognitieve processen spelen een belangrijke rol in het menselijk handelen en ook bij de totstandkoming van oordelen. Deze processen stellen ons in staat om grote hoeveelheden informatie te verwerken, te prioriteren en te groeperen, en op basis hiervan conclusies te trekken, ook wanneer die informatie soms ambigu en incompleet

is.⁶ Juist deze processen hebben echter ook specifieke kwetsbaarheden tot gevolg die onder meer kunnen leiden tot selectiviteit en vertekening van informatie.

Een voorbeeld hiervan in de forensische context is het zogenoemde ‘contexteffect’^{7,8,9}. Deze term geeft aan dat de resultaten van een forensisch onderzoek beïnvloed kunnen worden door de omstandigheden waarin het onderzoek wordt uitgevoerd, en in het bijzonder door de informatie die bekend is bij de forensisch onderzoeker die belast is met het onderzoek.¹⁰ Regelmatig wordt in de forensische literatuur ook de term ‘confirmation bias’ gehanteerd. Globaal kunnen er drie soorten van zulke informatie worden onderscheiden.

Allereerst is er de ‘base rate’ informatie die los staat van het specifieke onderzoek maar die een effect kan hebben op de verwachtingen van de onderzoeker. Het feit dat in het verleden bijvoorbeeld 95 van de 100 documenten vervalst bleken, kan van invloed zijn. De onderzoeker kent mogelijk meer waarde toe aan informatie die de verwachte conclusie ondersteunt. Die invloed is echter ongewenst omdat deze informatie in de bewijskracht van het deskundigenonderzoek zelf geen rol hoort te spelen.

Vervolgens is er de domein-irrelevante zaakinformatie. Zo kan het feit dat een verdachte bijvoorbeeld al drie maal eerder voor een soortgelijk vergrijp is gearresteerd en ook in deze zaak heeft bekend een document te hebben vervalst, relevant zijn voor een rechter, maar is het voor de handschriftonderzoeker volstrekt domein-irrelevante informatie. Al zal een professionele onderzoeker niet automatisch meegaan met de suggestie, en kan het ook het tegenovergestelde effect bewerkstelligen, hij zou niet moeten worden blootgesteld aan deze informatie. In een experiment van Dror *et al*¹¹ kregen vijf deskundigen een vingerspoot en vingerafdruk voorgelegd. Ze hadden die al eens eerder beoordeeld in een echte zaak, waarbij ze hadden geconcludeerd dat ze van dezelfde vinger afkomstig waren. Dit maal werden vingerspoot en vingerafdruk

-
6. Dror, I. E. (2009). How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics: The Journal of Law, Science, and Technology*, 50, 93-110.
 7. Saks, M.J., Risinger, M.D., Rosenthal, R. & Thompson, W.C. (2003). Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States, *Science and Justice*, 43, 77-90.
 8. Risinger, M.D., Saks, M.J., Thompson, W.C. & Rosenthal, R. (2002). The Daubert/Kumbo implications of observer effects in forensic science: hidden problems of expectation and suggestion, *California Law Review*, 90, 1-56.
 9. Broeders, A.P.A. (2009). De blinde onderzoeker, *Trema*, 6, 237-243.
 10. 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.
 11. Dror, I. E., Charlton, D. & Peron, A. (2006) Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156, 74-78.

aangeboden alsof het een nieuwe zaak betrof, en in een context die sterk suggereerde dat ze niet van dezelfde vinger afkomstig zouden zijn. Vier van de vijf deskundigen trokken vervolgens een andere conclusie dan ze eerder hadden getrokken op basis van hetzelfde spoor en afdruk.

Een derde vorm van beïnvloeding kan schuilen in de manier waarop het onderzoek wordt uitgevoerd. Als het betwiste materiaal gelijktijdig met het referentiemateriaal wordt onderzocht, dan kan het effect van ‘post hoc target shifting’ optreden. Van ‘post hoc target shifting’ is sprake als de relevant geachte aspecten van het betwiste materiaal mede worden bepaald door wat de onderzoeker heeft gezien in het referentiemateriaal. Thompson¹² liet bijvoorbeeld zien dat kennis van het DNA-profiel van de verdachte van invloed kan zijn op de interpretatie van een betwist (onvolledig en complex) DNA-profiel. Door die kennis ontstaat het gevaar dat de deskundige kenmerken van het referentieprofiel in het spoorprofiel ‘gaat zien’ die hij anders niet gezien zou hebben.

Het moge duidelijk zijn dat dergelijke vormen van beïnvloeding de juistheid van conclusies in gevaar zouden brengen doordat de conclusie niet zuiver is gebaseerd op het van toepassing zijnde ‘bewijs’. De vraag die rijst, is natuurlijk in hoeverre deze effecten daadwerkelijk optreden in de praktijk. Dror en Cole¹³ geven een overzicht van de studies naar contexteffecten in zaakonderzoek. De weinige studies die zijn verricht door forensisch wetenschappers en gedragswetenschappers spreken elkaar gedeeltelijk tegen. Hall en Player¹⁴ bijvoorbeeld, komen tot de conclusie dat het allemaal wel meevalt. Terwijl andere studies wel degelijk effecten laten zien.

Op voorhand is niet duidelijk of er wel of geen contexteffect zal optreden in een specifieke zaak. Maar zou het daarom niet het beste zijn om daar waar haalbaar het forensisch onderzoek zo in te richten dat de kans op ongewenste beïnvloeding wordt geminimaliseerd?

-
12. 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; zie ook: Meulenbroek, A.J., Kloosterman, A.D. & de Blaeij, T.J.P. (2009). Richtlijnen borgen onbevooroordeeld DNA-onderzoek. *Expertise en recht*, 2009, 119-129.
 13. 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, 161-167.
 14. Hall, L.J. & Player, E. (2008). Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Science International*, 181, 36-39.

Oplossingen

De methoden die contexteffecten kunnen voorkomen of verkleinen hangen af van het type informatie dat aan die effecten ten grondslag ligt.

Het effect van domein-irrelevante zaakinformatie kan worden aangepakt door zo veel mogelijk niet relevante zaakinformatie weg te houden bij de deskundige die het onderzoek uitvoert. Dit is ten dele mogelijk door het instellen van een getrapte procedure waarbij de deskundige die de zaak aanneemt en coördineert niet dezelfde is als de deskundige die het daadwerkelijke onderzoek uitvoert. Goed opgezette procedures kunnen zo een groot deel van de informatie weghouden bij de onderzoeker, en staan in principe los van het daadwerkelijk uit te voeren onderzoek.

Helaas zal een dergelijke procedure soms niet alle domein-irrelevante informatie weg kunnen houden. De context kan namelijk blijken uit de inhoud van het betwiste materiaal zelf, zoals vaak voorkomt bij gebieden als handschrift- en spraakonderzoek. Aanpassing van de onderzoeksmethode is dan noodzakelijk om de onderzoeker 'blind' te houden voor de context van de zaak. Zo zal bij het forensische handschriftonderzoek soms een selectie gemaakt moeten worden uit het aangeboden betwiste materiaal opdat de deskundige uit de geschreven teksten niet kan opmaken wat de context is van de zaak.

Daarnaast zou de deskundige in eerste instantie ook geen kennis moeten hebben van het referentiemateriaal in de zaak ter voorkoming van 'post hoc target shifting'. Böttcher¹⁵ en Froentjes¹⁶ stelden al in de vijftiger en zestiger jaren van de vorige eeuw dat het betwiste materiaal eerst onderzocht moet worden, en pas daarna het referentie materiaal. Dat wat de deskundige ziet in het betwiste materiaal, mag niet beïnvloed zijn door wat hij heeft gezien in het referentiemateriaal. De overkoepelende term voor het weghouden van dit type informatie is 'blind onderzoek', en van een goed uitgevoerd blind onderzoek mag men verwachten dat het deze contexteffecten tot een minimum beperkt.

Blind onderzoek voorkomt echter niet het effect van base rate informatie, omdat dit los staat van de context van de zaak. Een mogelijke oplossing hiervoor is het toevoegen van een (groot) aantal 'fakezaken' waarin de correcte conclusie juist niet de tot nog toe meestal correcte conclusie is. Hoewel effectief, zal duidelijk zijn

-
15. Böttcher, C.J.F. (1954). Theorie en praktijk van de gerechtelijke schriftvergelijking, *Tijdschrift voor Strafrecht*, LXIII, 77-131.
 16. Froentjes, W. (1969). Schriftonderzoek en statistiek, *Nederlands Juristenblad*, No. 33, 821.

dat deze oplossing niet erg werkbaar is, omdat het lastig kan zijn om veel realistische zaken te creëren waarbij de deskundige daadwerkelijk niet weet dat het een fake zaak betreft. Momenteel loopt bij het NFI in samenwerking met de Universiteit van Amsterdam een dergelijk onderzoek bij het deskundigheidsgebied Wapens en Munitie. Gedurende de komende twee jaar zullen op onbekende tijdstippen zaken worden aangeboden waarvan pas achteraf bekend zal worden dat het een fakezaak betreft. De resultaten van dit onderzoek zullen tevens een inschatting mogelijk maken van hoe haalbaar de toevoeging van fakezaken is, en of men erin slaagt realistische zaken te creëren.

In de literatuur wordt met enige regelmaat de (blinde) evidence line-up aangedragen als remedie tegen de vertroebelende effecten van de context. In het stuk van Merckelbach² en een stuk van Broeders⁹ wordt de evidence line-up ook aangedragen als de remedie tegen ongewenste invloeden van irrelevante informatie op het deskundigenoordeel. Oppervlakkig beschouwd lijkt de evidence line-up inderdaad veel te bieden, maar nadere beschouwing wijst anders uit. Veel van de adviezen gericht op gebruik van de evidence line-up ter voorkoming van de vertroebelende effecten van ondermeer a priori verwachtingen en context zijn terug te voeren op de gezaghebbende publicatie van Risinger *et al.*,⁸ maar al in 1984 en 1987 publiceerde Miller¹⁷ over de toepassing ervan bij handschrift- en haaronderzoek. Veel mensen benadrukten na Risinger *et al.* het belang van de evidence line-up zonder daadwerkelijk in te gaan op hoe deze dan op te zetten en uit te voeren, en wat het resultaat daarvan zou zijn.

Het gebruik ervan lijkt met name gebaseerd te zijn op de hoge indrukvaliditeit, wat wil zeggen dat de methode valide *lijkt*. Hoge indrukvaliditeit van een methode is een prettige, maar niet noodzakelijke eigenschap van een methode. Het behoort bovendien tot de meest subjectieve vormen van validiteit die bestaan, en biedt geen enkele garantie voor feitelijke validiteit. Empirisch onderzoek naar de evidence line-up is zo goed als afwezig. Een grondige analyse van de evidence line-up lijkt ons echter te leren dat het nut ervan in zaakonderzoek betrekkelijk klein is. Risinger *et al.* lijken zich hier wel van bewust, in tegenstelling tot Merckelbach en Broeders, waar ze schrijven⁸ (p. 49-50):

17. Miller, L. (1984). Bias among forensic document examiners: A need for procedural changes. *Journal of police science and administration*, 12, 407-411; Miller, L. (1987). Procedural Bias in Forensic Science Examinations of Human Hair, *Law & Human Behavior*, 157, 159-162.

‘Proper evidence line-ups present some nontrivial problems of design...

This process would obviously be easier for some types of examinations than for others. Unfortunately, it may often be most difficult precisely where it is most needed, in those areas, such as handwriting identification, with the least instrumentation and greatest subjectivity.’

Blind onderzoek en de evidence line-up

Alvorens we de beperkingen van de evidence line-up nader toelichten zullen we eerst een onderscheid maken tussen blind onderzoek en de evidence line-up. Met een blinde onderzoeksprocedure wordt een procedure bedoeld waarbij de deskundige bij de beoordeling van het betwiste materiaal geen kennis heeft van de domein-irrelevante zaakinformatie of van het referentiemateriaal. Een evidence line-up, daarentegen, is een procedure waarbij de deskundige diverse items krijgt gepresenteerd waarbij het hem niet bekend is wie of wat de bron is van de aangeboden items.

Er kunnen vervolgens twee soorten line-ups worden onderscheiden. Ten eerste de standaard line-up waarbij het betwiste materiaal wordt vergeleken met een line-up van referentiemateriaal. De deskundige weet dus wat het betwiste materiaal is, en moet dit vergelijken met meerdere andere items waarvan er enkele afkomstige zullen zijn van een verdachte, en meerdere andere toegevoegde items die in een bepaalde mate lijken op het betwiste materiaal. Deze toegevoegde items worden ook wel ‘fillers’ genoemd. Dit is de variant waar in de Engelstalige literatuur over is geschreven, en deze lijkt het meest op de zogenaamde Oslo-confrontatie, oftewel de ooggetuigeconfrontatie.

Broeders heeft daarnaast een variant ontwikkeld waarbij ook het betwiste materiaal wordt opgenomen in de line-up. De deskundige weet dan zelfs niet wat het betwiste materiaal is. Deze variant lijkt dus eerder op een soort blinde clustertaak voor de deskundige. Deze variant wordt zo nu en dan ook op het NFI toegepast bij het spraakonderzoek, maar dan met name om de deskundige de (subjectieve) overtuiging te geven dat hij, al dan niet, iets met het materiaal kan.

Het lijkt allemaal zo eenvoudig. De deskundige verzoekt om naast het betwiste schrift en het referentiemateriaal van X (a en b in het stuk van Merckelbach) ook de

beschikking te krijgen over een groep c van nagebootste handschriften en handtekeningen van X:

‘De documenten uit a, b en c worden als willekeurige serie aan de deskundige aangeboden en vervolgens speurt hij naar overeenkomsten en verschillen. Zo’n opzet levert het begin van een zorgvuldige test. Ze laat toe dat men de expert kan betrappen op fouten. Als de expert demonstreert dat hij foutloos door de documenten heen kan loodsen, pleit dat voor zijn expertise’

(Merckelbach,² p. 416).

Voor de keuze van de documenten in groep c (de fillers) zijn echter geen criteria voorhanden, terwijl deze keuze in hoge mate bepalend is voor de resultaten van het onderzoek.

Gebruik je bijvoorbeeld willekeurige handschriften als fillers in de line-up, dan zal dat weinig aan het onderzoek toevoegen. De line-up is dan vooral een wassen neus die gericht lijkt op de eerder genoemde ‘indruksvaliditeit’. Laat je iemand handschriften nabootsen, zoals Merckelbach voorstelt, dan maak je het de onderzoeker al wat lastiger, maar dan test je vooral ook hoe goed de nabootser is in het nabootsen van handschriften. Terwijl het feitelijk alleen maar gaat om de kracht van het bewijs.

De meest interessante optie is wellicht om een tweede deskundige de best gelijkende handschriften uit te laten zoeken uit een grote verzameling van handschriften. Maar ook dan test je meerdere dingen tegelijk: hoe goed doet de eerste deskundige het ten opzichte van de tweede, en hoe groot is de verzameling waaruit hij kan kiezen.

Het is dus niet duidelijk dat een evidence line-up enige garantie biedt voor een betere inschatting van de bewijskracht van het bewijsmateriaal. De evidence line-up is een poging meerdere problemen tegelijkertijd op te lossen, maar faalt doordat er te veel variabelen tegelijk variëren. Testen van bewijsmateriaal en onderzoeker kan beter los van elkaar gebeuren.

Conclusie

Handschriftonderzoek is een gecompliceerde vorm van vergelijkend forensisch onderzoek, met een belangrijke subjectieve component. Doel daarbij is niet zozeer te zeggen wie het betwiste handschrift schreef, maar te beschrijven wat de waarnemingen aan het bewijsmateriaal waard zijn bij het beantwoorden van die vraag. Het onderliggende basisprincipe is niet dat een handschriftonderzoeker in staat zou zijn om een handschrift aan een unieke bron te koppelen ('discernable uniqueness'). Het onderzoek is er op gericht tot een diagnostische waarde te komen (ook wel likelihood ratio, of bewijskracht). Dit met dezelfde gedegen wetenschappelijke methodologie die in andere vergelijkende forensische onderzoeken wordt toegepast, zoals DNA-onderzoek.

Een belangrijk verschil is echter gelegen in de moeilijkheid de kenmerken in handschriften objectief te definiëren en te vergelijken. Dat maakt de wetenschappelijke basis van het onderzoek niet wankel, maar dat kan wel tot gevoeligheid voor ongewenste beïnvloeding leiden. Afhankelijk van het type beïnvloedende informatie bestaan er oplossingen die zulke beïnvloeding tegengaan. De evidence line-up is daarbij misschien een aansprekende, maar bepaald niet de meest aangewezen methode vanwege de fundamentele problemen die wij in dit artikel besproken hebben. Wij zijn daarbij uiteraard voor een wetenschappelijke aanpak, en niet een voor-wetenschappelijke aanpak.