Analysing decision logs to understand decision-making in serious crime investigations

Article (Accepted Version)

Running head: Analysing investigative decision logs

Analysing decision logs to understand decision-making in serious crime investigations

Coral J. Dando*
Dept. of Psychology
University of Westminster
London, UK
W1W 6UW

Thomas C. Ormerod
School of Psychology
University of Sussex
Falmer, UK
BN1 9QH

Manuscript type: Research Article

Word Count: 5587

Email: c.dando@westminster.ac.uk

Pre-publication copy: Accepted for publication in Human Factors : The Journal of the Human Factors and Ergonomics Society 21 July 2017

Abstract

Objective: To study decision-making by detectives when investigating serious crime through the examination of Decision Logs to explore hypothesis generation and evidence selection.

Background: Decision logs are used to record and justify decisions made during serious crime investigations. The complexity of investigative decision-making is well documented, as are the errors associated with miscarriages of justice and inquests. The use of decision logs has not been the subject of an empirical investigation, yet they offer an important window into the nature of investigative decision-making in dynamic, time-critical environments.

Method: A sample of decision logs from British police forces was analyzed qualitatively and quantitatively to explore hypothesis generation and evidence selection by police detectives.

Results: Analyses revealed diversity in documentation of decisions that did not correlate with case type, and identified significant limitations of the decision log approach to supporting investigative decision-making. Differences emerged between experienced and less experienced officers' decision log records in exploration of alternative hypotheses, generation of hypotheses, and sources of evidential enquiry opened over phase of investigation.

Conclusion: The practical use of decision logs is highly constrained by their format and context of use. Despite this, decision log records suggest that experienced detectives display strategic decision-making to avoid confirmation and satisficing that affect less experienced detectives.

Application. Potential applications of this research include both training in case documentation and the development of new decision log media that encourage

detectives, irrespective of experience, to generate multiple hypotheses and optimize the timely selection of evidence to test them.

Key Words: Decision Logs; Crime Investigation; Heuristics & Biases; Hypothesis Generation; Expertise.

Precis: Decision Logs from British police forces were analyzed to explore hypothesis generation and evidence selection by senior detectives. Significant limitations of the decision log approach to supporting investigative decision-making emerged. There were differences between experienced and less experienced officers' use of decision logs for triggering the generation and testing of hypotheses.

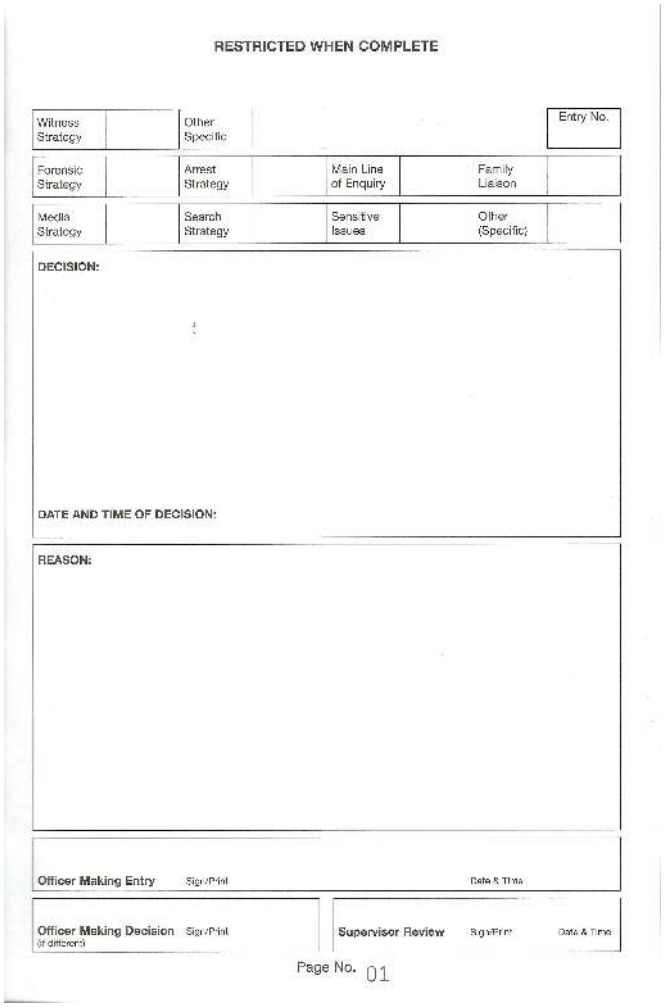
Introduction

Police decision making is under-researched, and so is not well understood. One starting point is to look at the records they make during investigations. Police detectives in the United Kingdom are accountable for their decisions and have to provide a mandatory record of what they did in sequentially numbered books called 'Decision Logs' (e.g. see Figure 1). These are auditable, hard-copy documents used to record, justify, and share decisions made during serious crime investigations (ACPO Crime Committee, 1999; College of Policing, 2014). Each decision is entered on a separate page and every decision is timed, dated, and signed by the officer making the decision. Although guides to best practice exist (e.g., the UK 'Investigative Doctrine' - ACPO, 2006), there is no gold standard against which to compare performance. The UK College of Policing (the professional body for policing in the UK) has recently adopted a National Decision Model (NDM) '*to help everyone in policing make decisions*'. NDM is descriptive and procedural, comprising six key elements to be considered when making *all* decisions. However, NDM does not specifically encourage the generation and testing of hypotheses, and so is likely to result in a preponderance of procedural decisions (i.e., formulaic decisions that follow expected practice). While prescription is, to some extent, unavoidable, the need to generate and test alternative hypotheses is also important to the investigative process.

The complexity of investigative decision-making is widely recognised (e.g., Alison et al., 2014; Eyre & Alison, 2007; Schulenberg, 2014; van den Heuvel, Alison, & Power, 2014; Vickers & Lewinski, 2012), as are errors emerging from miscarriages of justice and serious case reviews (e.g., Ellison & Morgan, 2015; Leo, 2008). Recently, the UK Home Affairs Select Committee (2011) raised concerns over decisions made during public order incidents in August 2011, and a serious case

review concerning the murder of a teenage girl revealed erroneous decision-making and decision avoidance (West Mercia Police, 2015). Here, we report a study of decision logs, which reveals significant limitations of decision logs for supporting crime investigation. The study also explored the presence of biases in decision log records that have been shown in other domains to affect hypothesis generation and testing.

Figure 1. Example of a decision log


RESTRICTED WHEN COMPLETE

Witness Strategy		Other Specific				Entry No.
Forensic Strategy		Arrest Strategy		Main Line of Enquiry	Family Liaison	
Media Strategy		Search Strategy		Sensitive Issues	Other (Specific)	

DECISION:

DATE AND TIME OF DECISION:

REASON:

Officer Making Entry Sign/Print Date & Time

Officer Making Decision (if different) Sign/Print

Supervisor Review Sign/Print Date & Time

Page No. 01

The Nature of Investigative Decision-Making

A simple characterization of investigating serious crimes is as a task with two components: hypothesis generation (e.g., determining modus operandi, identifying

suspects), and hypothesis testing (e.g., seeking evidence concerning crime scenes, alibis and other sources). Decision-making tasks such as these can be subject to the use of cognitive heuristics that are known to cause biases in responses (e.g., Tversky & Kahneman, 1973), and it is the impacts of these biases that concern us here.

One such heuristic is 'satisficing' (Simon, 1956, 1990), where individuals limit the space of possible ideas that must be searched for a solution by generating a single solution idea that is satisfactory and suffices (hence 'satisficing') to meet the current goal. This reduces cognitive load, but may not give the optimal solution. Theoretical analyses typically suggest that an optimal approach to hypothesis generation is to conduct an exhaustive search for as many hypotheses as possible (e.g., King et al., 2004). As noted in the ACPO (2006) Investigative Doctrine, investigating officers should consider all possible explanations for any crime or evidence set. In the domain of investigation, the effect of satisficing is to limit the hypotheses generated by investigators, typically to those that most obviously or immediately explain the available evidence. The effect of satisficing, therefore, goes against the prescriptively optimal approach of generating alternative hypotheses as exhaustively as possible.

Evidence for satisficing can be found in a range of domains. For example, automobile mechanics, irrespective of expertise, were found to generate fewer than one fifth of possible hypotheses, despite being confident their explanations were exhaustive (Mehle, 1982). In an investigative domain, Fahsing and Ask (2016) found that police officers generated only 50% of the hypotheses subsequently identified as representing a gold standard for each case they examined. Here, the 'gold standard' comprised all the hypotheses that should be considered for any specific evidence set for a presented case, and was established by a panel of senior police investigators.

The impact of satisficing on investigative hypothesis generation has been found to be affected by expertise, but not always in a straightforward way. Wright (2013) found that inexperienced UK police officers were more likely than experienced officers to fixate on single explanations of crimes, yet Fahsing and Ask (2016) found the opposite result with Norwegian police officers. Alison and colleagues (2013) reported that perceived time pressure rather than experience reduced the generation of investigative hypotheses. Sandham (2013) found that both inexperienced and experienced police officers failed to generate all possible hypotheses consistent with the presence of a piece of evidence whose validity was uncertain, and were more likely to generate hypotheses consistent with the guilt of a person of interest. Her results are consistent with truth and lie response biases typically found with general public and law enforcement participant groups, respectively). Truth bias is a default position adopted whereby people tend to believe accounts of others, whereas law enforcement officers have a tendency to disbelieve what they are told. (e.g., Meissener, & Kasin, 2002; Masip, Garrido, & Herrero, 2009)

Just as cognitive heuristics can affect hypothesis generation, the biases they produce are also evident in hypothesis testing. The prescriptively optimal approach to hypothesis testing is agreed to be hypothetico-deductive falsification (e.g., Tarantola, 2006; Magee, 2013), in which evidence is sampled to try to disconfirm the current hypothesis, the corollary being a failure to disconfirm provides corroborative support. However, empirical studies suggest that individuals demonstrate ‘confirmation’ bias (Wason, 1966): a tendency to seek or accept evidence supporting the current hypothesis. Ask and Granhag (2005) found both naïve individuals and law enforcement personnel showed confirmation bias when sampling evidence to test

hypotheses, but experienced investigators were affected by a guilt bias, an involuntary or automatic tendency to assume guilt (Kassin, Goldstein, & Savitsky, 2001; Meissner & Kassin, 2002). The effect of guilt bias was to reduce the impact of confirmation bias on hypothesis testing, where confirming evidence might exonerate the person of interest. Confirmation and guilt biases may occur because of an overarching 'availability' bias (Tversky & Kahneman, 1973), in which investigators make decisions based on how easily examples from previous experience come to mind. In medical diagnosis, the order in which pieces of evidence are presented influences final diagnosis, with early disease-indicative evidence dominating decisions even when undermined by later evidence (Chapman, Bergus, & Elstein, 1996; Rebitschek, Krems, & Jahn, 2015). Like confirmation bias, order effects arising through availability can impair the sampling of evidence to test investigative hypotheses.

Empirical evidence for biases in hypothesis generation and testing typically comes from laboratory studies. However, naturalistic methods sometimes refute the presence of systematic bias in performance. For example, Hutchins (1995) found individual's overconfidence bias all but disappears in collaborative task performance. Mossmann (2013) investigated the decisions made by forensic examiners and reported random decision making errors rather than systematic bias. Ball, Maskill, and Ormerod (1998) found little evidence for satisficing strategies in idea generation behaviours of experienced designers. Likewise, experienced insurance fraud investigators pursued multiple hypotheses in parallel (Ormerod, Barrett, & Taylor, 2008) as did doctors when making diagnostic decisions (Alby, Zucchermaglio, & Baruzzo, 2015). Decision making in natural settings can differ markedly from typical laboratory research because it rarely occurs in sanitized contexts, and is often

mediated by factors such as colleagues/team members and technology (Blumenthal-Barby & Krieger, 2015).

We examined decision logs to understand whether crime investigators reveal satisficing and confirmation biases in their records. We summarised different types of log entry, looking at how decision log structure interacts with the nature of the crime, and how log characteristics vary across individuals, and as a function of investigative experience. We then analysed a set of case exemplars. Finally, we explored records of generation and testing of investigative hypotheses and evidence, examining whether there was evidence for satisficing and confirmation biases.

Methods

Summarisation and Data Reduction

Sixty decision logs were randomly selected from the repositories of two UK police forces blind to the research aims. The authors worked independently to identify entries as 'decisions' using the following criteria, which all had to be present: i) entries concerned the crime itself, ii) the detective had made clear a preference of possible action, and iii) a reason was given to follow the course of action. Twelve randomly selected decision logs (20%) were passed to two independent researchers for recoding. Inter-rater reliability, assessed for each decision log independently by comparing codes supplied by each rater to each entry (decision; not decision), revealed highly significant levels of agreement for the number of decisions in all logs, all Kappas > .935, all *p*s < .001.

Exploration of Investigative Decisions

We conducted a detailed exploration of the timeline of investigative decision-making in the logs, illustrating key recurring themes with reference to three case exemplars, changing nothing in the reported decisions except to ensure anonymity.

137 We drew case timelines plotting the generation and testing of hypotheses against

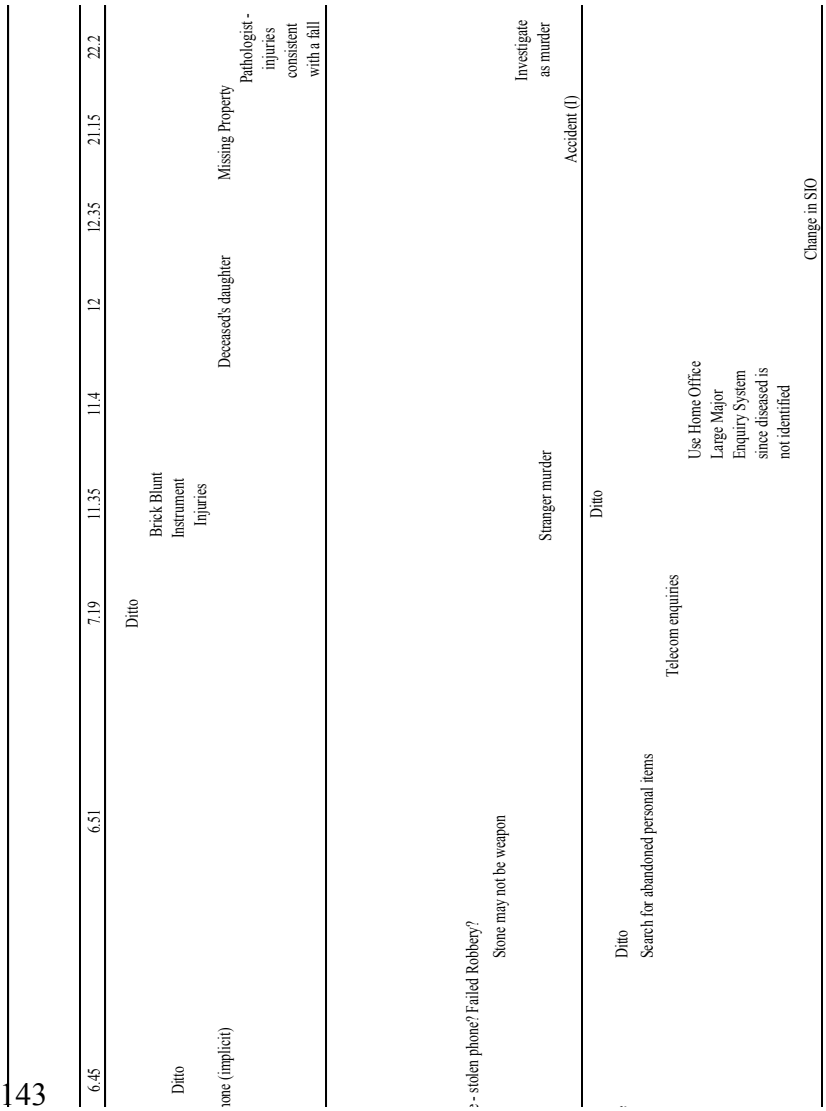
138 evidence collection over time (Table 1).

139

140 Table 1.

141 *Case timelines plotting the generation and testing of hypotheses against evidence*

142 *collection over time for Case Study 2: Stranger murder.*



143

Using a Grounded Theory approach (e.g., Charmaz & Henwood, 2007), we identified key moments in a decision log where the course of an investigation changed ('tipping points', according to Fahsing & Ask, 2016). We examined these points for recurrent behaviours associated with hypothesis generation and evidence selection.

Counts of Hypothesis Generation and Testing

Logs were then examined to identify the numbers of distinct hypotheses generated, the amount of evidence sources examined in order to test these hypotheses, and the order in which they were generated. These counts were taken from a representation of the hypotheses and evidence referred to in each log using problem behaviour graphs (Ericsson & Simon, 1993), in which top-level hypotheses are considered as problem goals and sub-hypotheses that relate to the top-level hypothesis are connected by branches from this node. Representing hypotheses as a problem behaviour graph enables a definition of rules as to what determines a new hypothesis distinct from previously generated ones, and facilitates tracking of hypothesis generation and testing. Figure 2 illustrates a problem behaviour graph for the 'Stranger murder' described below (Case Study 2).

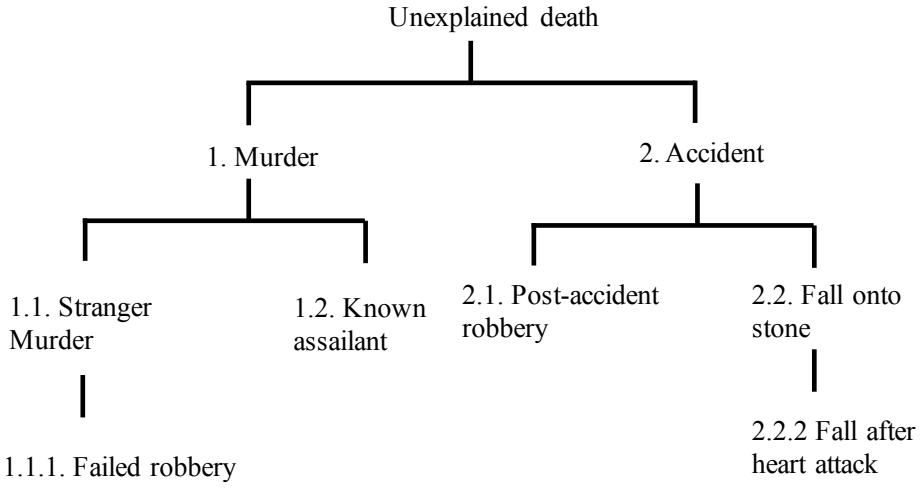
Once the first hypothesis is created, distinct hypotheses are either added at the same level in a breadth-first expansion of the graph, or as variants of that initial hypothesis in a depth-first expansion of the graph. Thus, we defined a hypothesis as a distinct addition to the graph under the following coding rules:

1. If it established a new line of investigation. For instance, "The victim was murdered" and "The victim suffered accidental death" are distinct hypotheses at the same level;
2. If it modified an existing hypothesis with a new line of enquiry. For example, if a previously mentioned hypothesis was "The victim was assaulted by an

unknown is a distinct hypothesis below the level of the hypothesis "The victim was murdered" and at the same level as the previously mentioned hypothesis;

3. If it extended an existing hypothesis with a more focused line of enquiry. For example, if a previous hypothesis was "the victim was assaulted by an unknown assailant", and a new hypothesis stated "the victim was assaulted by unknown male assailant", then the latter became a new node at a level below the previous hypothesis.
4. Counts were made of the number of entries in a decision log showing transitions horizontally or vertically between hypotheses, following the method of Ball & Ormerod (1995). A ratio of horizontal to vertical transitions greater than 1 indicates consideration of multiple alternative hypotheses in parallel, while a ratio less than 1 indicates satisficing behaviour.

Figure 2. Hypothesis generation graph (example from Case Study 2). The top level hypothesis "Unexplained death" has two alternative hypotheses in the decision log: 1. Murder and 2. Accident. Each of these in turn has a number of hypotheses associated with it.



Strategies for hypothesis generation and testing are likely to change over time, since different phases of an investigation yield different amounts of evidence and investigative activity. To examine whether generation of hypotheses, opening of evidence sources, and activity transitions varied over time, we counted these items across four quartiles, each containing 25% of the log entries for each case. We divided quartiles by number of entries rather than time because the time-course of investigations is highly variable, and affected by non-investigatively relevant factors (e.g. staff availability, courts processes, delays in evidence processing). In choosing entry counts as a metric for sectioning the logs, we aimed to capture the fact that all investigations will have initial and end phases with at least one interim phase.

In addition, we examined whether the number of years of experience in leading investigations would impact the use of decision logs. Seven officers had experience of five years or more (M = 10.40 years, ranging from 5 to 16 years), while the remainder (7) had experience of three years or less (M = 2.00 years, ranging from 1 to 3 years). Thus, data analysed were the average numbers of hypotheses generated, evidence sources opened, and activity transitions made by each SIO in each quartile averaged across cases in which they were involved.

Results

Case Summarization

Table 2 shows the total number of decision logs and crime types, and mean number of decisions and SIOs. A multinomial logistic regression was conducted to examine whether case type predicted number of investigation days, number of log entries or number of SIOs, but the model was not significant, $\chi^2(4) = 0.91$, $p = .412$. The mean number of entries made for each week of a case by experienced investigators ($M_{entries} = 8.19$, SD = 4.13) and less experience investigators ($M_{entries}$ =

213 9.62, SD = 3.30) did not differ significantly, $t = 1.14$, $p = .445$, $d = .31$. Nor did the
214 mean number of words per entry ($M_{\text{experienced SIO words}} = 36.62$, SD = 21.12; $M_{\text{less experienced SIO words}} = 29.59$, SD = 23.50), $t < 1$.
215
216
217 Table 2.
218 *Total number of decision logs and crime types, and mean number of decisions and*
219 *senior investigating officers.*

Crime Type	Number of Logs Analyzed*	Number of Investigation Days	Total Number of Log Entries	Number of SIOs
Murder	28	86.86 (SD = 61.03)	86.14 (SD = 34.24)	3.20 (SD = 1.23)
Aggravated Burglary	11	66.34 (SD = 23.54)	84.45 (SD = 87.21)	1.70 (SD = 0.41)
Sexual Offences	12	35.68 (SD = 12.34)	34.45 (SD = 14.30)	1.90 (SD = 1.12)
Arson	4	78.43 (SD = 23.24)	88.32 (SD = 101.65)	2.30 (SD = 2.42)
Other	5	101.43 (SD = 64.71)	122.40 (SD = 133.20)	4.43 (SD = 4.56)

* Each case has one continuous decision log

220
221 **Exploration of Investigative decisions**
222 A number of themes emerged across the cases, which can be divided into two
223 categories: modifiers of decision-log entry frequency and type, which we describe
224 with reference to the whole sample; and themes about hypothesis generation and
225 testing, which we illustrate with reference to three case studies.
226 One unexpected factor that appeared to increase duration and number of case
227 log entries was when a case raised major social and behavioural side-issues. As an
228 extreme example, our biggest case (200+ logs extending over three case booklets) was
229 an aggravated breach of an Anti-Social Behaviour Order involving two warring
230 families. Whereas murder enquiries tend to take longer than aggravated
231 burglary/Grievous Bodily Harm enquiries, the latter tended to have more entries

concerning social/behavioural issues (e.g., mental health, witness protection) and so generated much more variability in the number of decisions that were logged. In contrast, the sexual assault cases we looked at generated fewer log entries, but tended to involve unknown or unrelated assailants, which we suspect is not a particularly representative sample of sexual assault cases.

As one might expect, the average number of SIOs involved varied with case type and complexity, with murder enquiries typically having more SIOs than aggravated burglary. Sometimes SIOs changed due to availability (e.g., vacations), but sometimes were changed by tactical decisions made by commanding officers. Changes in SIO were frequently marked by a set of review logs, made as part of the handover. As case study 2 below illustrates, these change-over moments were often key change points in the direction of investigations.

Three case studies illustrate key themes in the decision logs concerning hypothesis generation and testing.

Case study 1: Drive-by murder. This case involved a revenge killing between gangs, which took place in a busy public place in broad daylight. A single SIO was assigned the case throughout the three-week investigation. Table 3 shows two log entries recorded at key moments in the investigation.

Table 3.

Decision log entries for Case study 1: shooting

Log Entry No	Time of Decision (Post notification of crime)	Decision	Rational
4	1 Hour 40 mins	Major Incident - Use Home Office Large Major Enquiry System	Given that the incident appears to be a 'stranger type' murder, Cat B, a significant amount of evidence is expected to be gathered…
24	24 Hours	At this time the motives for this death are unknown...Initial intelligence shows there is acrimony between Gang A to whom the victim belonged and Gang B. Approx. 2 months ago a tattoo parlour was targeted by arsonists…the tattoo parlour was the premises used by Gang B. Furthermore, there was a road rage attack (X days ago) on XXXX who was affiliated to Gang B	A number of hypotheses exist at this time: 1) non-discriminatory shooting by other XXXX, 2) non-discriminatory shooting by others not associated to the XXXX, 3) deliberate shooting of XXXX by XXXX or otherwise because of the victim's personal lifestyle, 4) deliberate shooting of XXXX by XXXX or otherwise because of his affiliation to XXXX believed to be Gang A

The first (entry no. 4) was made 1 hour 40 minutes after the incident was first reported. The initial hypothesis reported (that the incident is a drive-by shooting) turned out to be correct, and influenced the following 20 log entries, recorded over 24 hours. However, the next day, the detective documented his investigative strategy (entry no. 24), where he explored complexities surrounding the initial intelligence, which implicitly set up the consideration of motives for the shooting. This led him to flesh out different hypotheses that the investigation needed to entertain.

This generation of multiple hypotheses appears to alert the SIO to the importance of undertaking victimology research via the victim's partner and other associates, partly to rule out the possibility that the shooting was a result of something other than a revenge attack (hypothesis 3 shown in Log 24 allows that it is a deliberate

shooting by 'others' because of lifestyle, e.g., a personal relationship motive). Thus, the course of the investigation was influenced by widening the scope of evidence sought, and allowed collection of evidence to test the initial hypothesis of a revenge attack. Here we see how evidence can serve both confirmatory and disconfirmatory roles if selected appropriately. The SIO assigned this case was the most experienced in our sample (>16 years).

Case study 2: "Stranger murder". A man was found dead in a local park, with head injuries from a blunt instrument. Representing the case along a timeline reveals satisficing in the initial investigation. The case timeline shows initial consideration of a failed robbery, but once the idea was generated that this was a stranger murder (a general case of the failed robbery hypothesis), no other hypothesis was entertained for a considerable time. Even when a pathologist reported that wounds were consistent with a fall, generating an implicit hypothesis that it might be an accident, the only hypothesis that continued to be entertained was stranger murder. Indeed, the accident hypothesis was not stated explicitly in the log; instead the SIO made a note that the pathologist's contribution was unreliable and should be ignored. The logs to this point are consistent with the effect of a confirmation bias limiting the consideration of evidence that might pertain to alternative explanations of the incident.

A switch in SIO led to a change in investigative stance. The new SIO was relatively inexperienced (< 2 years), but had served under the SIO responsible for the successful drive-by shooting investigation. He introduced an immediate note of circumspection, illustrated by log 11, shown in Table 4. In log 20a, 21 hours after the incident, he explicitly states multiple hypotheses. In log 21, he notes, in stark contrast

to earlier investigation, that the cause of death is unknown. In fact, the final investigation outcome was of death by accident with no robbery having taken place.

Table 4.

Decision log entries for Case study 2: stranger murder

Log Entry No	Time of Decision (Post notification of crime)	Decision	Rational
11	10 Hours	Major Incident - Use Home Office Large Major Enquiry System	At this stage there has been no formal identification of the deceased, we have no suspects, and are uncertain of cause
20a	20 Hours	Mature Assessment' (where the facts are clear the SIO undertakes a mature assessment, assessing the broader range of investiagtive issues to determine the appropriate level of resources that are required from that time)	There are various hypotheses being considered: 1) this was a deliberate act …pushing the injured party onto a pointed object... being forced into his neck…part of a robbery; 2) the injured party fell on two occasions accounting for his injuries…property has been mislaid, not theft 3) the injured party fell on two occasions...he has had his property stolen from him when he was on the ground
21	21 Hours	Investigation to be conducted with the same resources at this time as a murder	The action to cause death is not clear …subject of a deliberate push or a fall

Case study 3: Disappearance. This case was the longest in the sample, lasting over two years, in which a woman initially reported missing by her husband became a murder enquiry. Investigators focused for nearly two years upon a single hypothesis, that the husband had killed and disposed of the victim's body. Although the hypothesis was in the end correct, the breakthrough in the investigation occurred only when an SIO re-evaluated evidence collected after the investigation had faltered with no action taken for nearly a year. A visit by UK police to the victim's country of

residence triggered a review of the evidence, which noted evidence pertaining to witnesses A1 and A2, shown in Table 5.

The recording of this evidence in the decision log (even though it had been available elsewhere for some time) is important, since it triggered a change in the investigation. In particular, the 'rationale' given in Log 27 contains a contradiction made explicit by recording it: why would the husband enquire about his wife's whereabouts and then tell them she had gone to see a friend who lived elsewhere in the country? This record triggered a declaration of the husband as a suspect, and is the 'information' referred to in Log entry 34 (see Table 5). The act of documenting information made the anomaly in the husband's behaviour more prominent, providing the first strong evidence of an inconsistency in his account.

Table 5.

Decision log entries for Case Study 3: Disappearance

Log Entry No	Time of Decision (Post notification of crime)	Decision	Rational
27	10 Months	Persons A1 & A2 to be treated as significant witnesses	A1 & A2 have significant information about the victim including a phone call made to them by XXXX enquiring into his wife's whereabouts and then telling them that she had gone to see a 'friend' in Benidorm
34	11 Months 2 weeks	XXXX to be declared a suspect…his arrest will take place when deemed appropriate	Information exists that demonstrates that spouse may be responsible for victim's disappearance/murder...

Analysis of Hypothesis Generation and Testing Counts

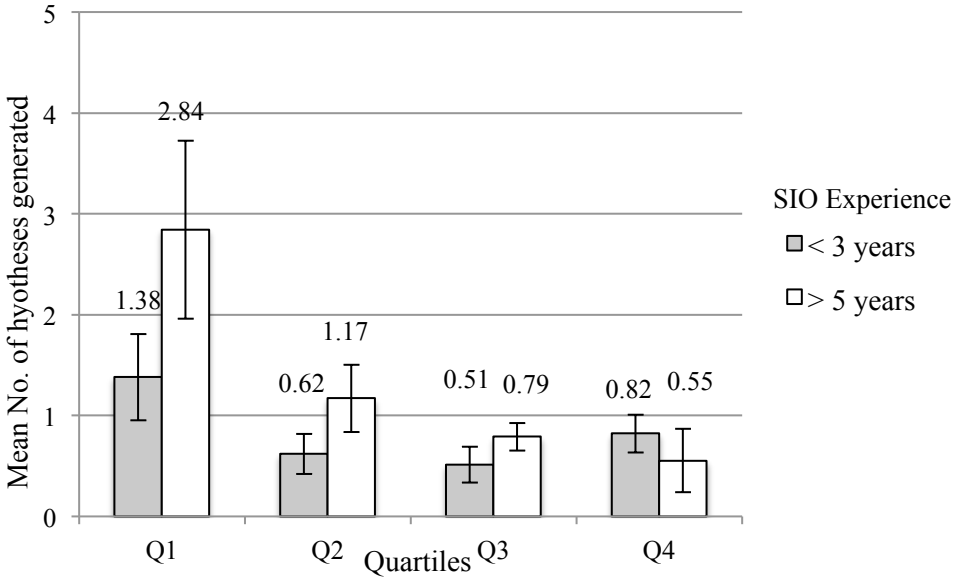
To investigate hypothesis generation and testing counts we conducted a series

of inferential statistical analyses as a function of experience, followed by post hoc t-test pairwise comparisons, applying Bonferroni correction.

A significant effect of quartile was found in hypothesis generation, $F(1.60, 19.25) = 25.53$, $p < .001$, $\eta^2 = .68$. More hypotheses were generated in quartile 1 ($M_{1st} = 2.11$, $SE = .25$; 95% CI [1.57, 2.66]), $p < .001$, than in quartiles 2 ($M_{2nd} = .89$, $SE = .10$; 95% CI [.68, 1.11]), $p < .001$, $d = .91$, 3 ($M_{3rd} = .65$, $SE = .06$; 95% CI [.53, .78]), $p < .001$, $d = .78$, and 4 ($M_{4th} = .69$, $SE = .09$; 95% CI [.48, .89]), $p < .001$, $d = .77$. No other pairwise comparisons were significant, all ps > .310.

There was a significant effect of experience, $F(1, 12) = 9.08$, $p = .011$, $\eta^2 = .43$. Experienced detectives documented more hypotheses ($M_{> 5 \text{ years}} = 1.34$, $SE = .12$; 95% CI [1.08, 1.60]) than less experienced ($M_{< 3 \text{ years}} = 0.83$, $SE = .12$; 95% CI [0.58, 1.09]), $p = .003$.

Figure 3. Mean hypotheses reported as a function of SIO experience (< 3 years; > 5 years) across decision log quartiles (bars show between subjects 95% confidence intervals).

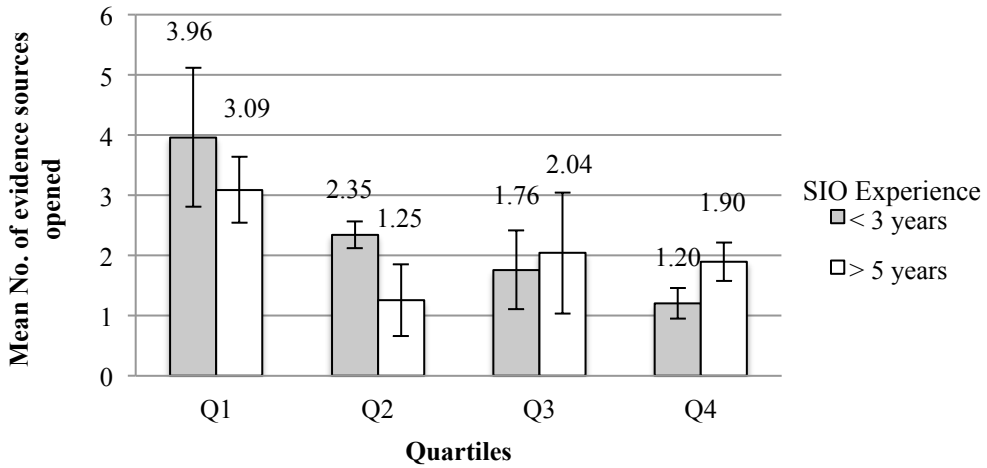


The experience X quartile interaction was significant, $F(1.60, 19.25) = 6.97$, $p = .008$, $\eta^2 = .37$. More hypotheses were documented by experienced than inexperienced detectives (see Fig. 3 above) in quartiles 1, $p = .011$, $d = .92$, and 2, $p = .038$, $d = 1.09$, with no significant difference between groups in quartiles 3 and 4, $ps > .215$.

Evidence Sources

A significant effect of quartile for evidence sources emerged, $F(1.95, 23.34) = 24.60$, $p < .001$, $\eta^2 = .67$. More evidence sources were opened in quartile 1 ($M_{1st} = 3.53$, $SE = .33$; 95% CI [2.82, 4.24]) than in quartiles 2 ($M_{2nd} = 1.80$ $SE = .16$; 95% CI [1.45, 2.15]), $p = .013$, $d = 1.11$, 3 ($M_{3rd} = 1.90$, $SE = .31$; 95% CI [1.23, 2.57]), $p = .011$, $d = .96$, and 4 ($M_{4th} = .1.55$, $SE = .10$; 95% CI [1.32, 1.77]), $p = .009$, $d = 1.01$. No other comparisons were significant, $ps > 0.411$. The main effect of experience was non-significant, $F < 1$.

Figure 4. Mean number of evidence sources opened as a function of SIO experience group (<3 years; > 5 years) across decision log quartiles (bars show between subjects 95% confidence intervals).

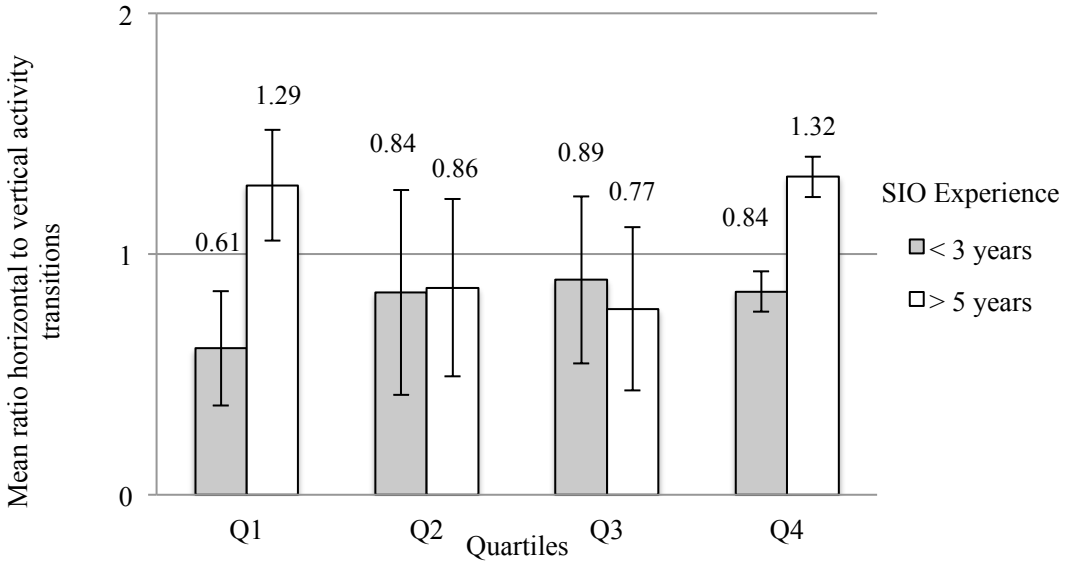


353 The interaction between experience and quartile was significant, $F(1.95, 23.34) = 5.72$, $p = .010$, $\eta^2 = .32$. More sources (see Fig. 4 above) were opened by less experienced detectives in quartiles 1, $p = .011$, $d = .2.23$, and 2, $p = .015$. $d = 1.09$, Experienced officers opened more sources in quartile 4, $p = .019$, $d = 2.11$, with no difference in quartile 3, $p = .712$. Less experienced officers sampled the evidence space more at the start of the investigation, while more experienced officers tended to sample towards the end of an investigation.

Vertical Activity Transitions

For horizontal to vertical activity transitions, the main effects of quartile, $F(3, 36) = 1.35$, $p < .274$, and experience, $F(1, 12) = 3.43$, $p = .090$, were non-significant. The quartile X experience interaction was significant, $F(3, 36) = 3.63$, $p = .02$, $\eta^2 = .23$.

Figure 5. Mean ratio of horizontal to vertical activity transitions as a function of experience group (<3 years; >5 years) across decision log quartiles (bars show between subjects 95% confidence intervals).



A larger ratio of horizontal to vertical activity transitions by experienced investigators emerged in quartiles 1, $p = .004$, $d = .84$, and 4, $p = .006$, $d = .91$, with no difference between groups in quartiles 2 and 3 (see Figure 5), $ps > .452$. Experienced officers switched across numerous hypotheses early and late suggesting a greater exploration of the hypothesis space, than less experienced officers.

Discussion

The summarization data indicate no clear relationship between decision log entries and factors such as crime type or duration of investigation. Detectives varied in the entries they made, some diligently documenting all hypotheses and evidence, others making scant records, but entries did not differ in frequency or length according to experience. This suggests that there are factors affecting the use of decision logs that reflect individual differences such as diligence and commitment to documentation. Despite being a legal requirement, there is clearly a large degree of discretion available to SIOs in the extent to which they document their thinking and decisions. However, some regularities are apparent in decision logs. Entries suggest that satisficing and confirmation biases do affect police investigations, but increasing expertise overcomes these biases to some extent. Experienced SIOs documented twice as many hypotheses as less experienced officers in the first two quartiles of decision logs.

Analysis of documented evidence sources also shows an effect of experience, Less experienced detectives documented more new evidence sources in quartiles 1 and 2 than more experienced detectives. Our interpretation of this finding, confirmed by inspection of the logs and the timelines for each case is that less experienced detectives tended to gather as much evidence as they could as quickly and as they could that corroborated a particular hypothesis. This behaviour is consistent with

confirmation bias, where multiple new evidence sources are pursued to corroborate a single hypothesis. We have previously suggested, however, that an aspect of investigative expertise is an ability to judge the right time to seek evidence (Ormerod et al., 2008). Indeed, there are instances where opening evidence sources too early appears to have hindered investigations. For example, an investigation into the Soham murders (https://en.wikipedia.org/wiki/Soham_murders), where school janitor Ian Huntley was eventually convinced of killing two schoolgirls, was significantly held up by the decision to issue a media call for information, which flooded the enquiry with false leads (Bichard, 2004).

Interestingly, experienced investigators documented more new evidence sources in the final quartile than less experienced investigators. In subsequent discussions, some experienced SIOs commented on using a tactic of 'withholding the obvious', that is, leaving some tests of a hypothesis until late into an investigation, as a final check prior to charging a person of interest with the crime. This behaviour is consistent with a disconfirmatory approach to hypothesis testing, in which a hypothesis is subjected to final challenge.

The analysis of transitions between hypotheses indicates less experienced detectives remained focused on single hypotheses. In contrast, in both the early and late phases of an investigation, more experienced investigators appear to have considered multiple hypotheses in parallel. The appearance early in an investigation of multiple alternative hypotheses suggests experienced investigators are aware of the benefits of keeping an open mind. Many studies have shown that experts tend to spend longer than novices on the problem understanding phase in tackling new problems (e.g., Runco, 1994).

The reduction in the transition ratios in quartiles 2 and 3 is consistent with following up of specific hypotheses, where specific lines of enquiry have been chosen as the focus of the ongoing investigation. A return to the consideration of multiple hypotheses in the later stages of an investigation may reflect the evaluative skills of experienced investigators who, in the process of evaluating a hypothesis before acting upon it, may return to previously dismissed explanations or search for new ones. Again, a test of this possibility requires fieldwork observations.

Externalisation, the process of moving knowledge or ideas from being stored internally in an individual's memory to an external environment such as a written, diagrammatic, pictorial or auditory form has been shown to aid cognition (e.g., Cox 1999). Externalisation can influence problem-solving and decision-making (e.g., Shirouzu, Miyake, & Masukawa, 2002; Steffensen, 2013). For example, fire and rescue incident commanders trained to explicitly verbalise thinking, increased their tendency to consider goals, consequences, and displayed enhanced situation awareness without an increase in response latency (Cohen-Hatton & Honey, 2015; Cohen-Hatton, Butler & Honey, 2015). Likewise, in higher education settings, when students working in dyadic settings were encouraged to verbalise multiple hypotheses, their task performance improved (Beckmann, Beckmann, Briney & Wood, 2015).

It appears from our analyses that externalisation also impacts upon criminal investigations, albeit that here externalisation was the process of completing the decision log. For example, in the drive-by shooting case, after 24 hours, the SIO documented his investigative strategy, in which he explored the complexities surrounding the initial intelligence and noted a number of alternative hypotheses that the investigation needed to entertain. A similar impact of externalisation, in this instance of the evidence held within the case, changed the course of the disappearance

444 investigation (Case Study 3). This type of externalised 'exploration' to flesh out
445 alternative hypotheses was also reported with experienced fraud investigators
446 (Ormerod et al., 2008). Here we also see an example of how evidence serves dual
447 confirmatory and disconfirmatory roles if selected appropriately, consistent with
448 Klayman and Ha's (1987) recasting of confirmation bias as a positive test strategy.
449 Our study has a number of key limitations, which make its findings
450 preliminary rather than definitive. We cannot know whether officers generated more
451 hypotheses than they actually documented. Accountability and self preservation may
452 have hindered the documentation of hypotheses (see Waring, Alison, Cunningham, &
453 Whitfield, 2013), or it may be that less experienced officers were simply more
454 cautious about documentation despite conceiving of multiple hypotheses so that they
455 would not appear uncertain or naïve. Yet, their training makes very clear that they
456 should both generate and document alterative hypotheses. Alternatively, they may
457 have documented fewer hypotheses because of the cognitive and time demands of
458 doing so, which might be better managed by more experienced officers. Individual
459 differences in time perceptions, rather than investigative experience, may also have
460 affected hypotheses generation, as has been reported in laboratory-based research
461 (Alison et al., 2013; Dougherty, Mathias, & Marsh, 2003). Distinguishing between
462 these explanations will require further research that studies decision-making
463 concurrently during ongoing investigations. Finally, although we asked the
464 collaborating police services to provide decision logs from a mixed but representative
465 a sample of cases, we cannot be sure that the sample was not biased by unknown
466 selection preferences. We are reasonably confident that this potential bias was not a
467 major concern, partly because of the wide range of cases covered, and partly because

in a number of instances the police services were not always represented in a positive light in the decision logs provided.

Practical Implications

The practical use of decision logs appears constrained by their format and context of use, arguably irreparably. In an environment where practice is constrained by legislation and legacy technology, it is difficult to see how decision logs can be used as collaborative decision support tools in an effective way. Replacing paper documents with online resources might overcome some of the problems, but it would not address the contextual limitation that SIOs may be cautious not to document anything that might negatively impact the prosecution case (e.g., ACPO, 1999; Tasca et al., 2012).

The generic, inflexible nature of decision logs is such that rather than supporting investigators to generate multiple alternative hypotheses, they appear to constrain hypothetical thinking by encouraging SIOs to first document each decision, and then provide a rationale. Externalizing is known to support cognition, and in dynamic investigative environments the pressure to make decisions is such that the benefits of multiple hypothesis generation may not be recognized, or simply overlooked, and the decision log format does nothing to mitigate this behaviour.

However, we found that experienced SIOs evidenced an ability to overcome biases in decision-making. Moreover, they documented their hypothetical thinking despite the decision log format, and were able to moderate biases in the decision-making of less experienced colleagues. This would suggest that if the format of decision logs was amended to encourage more effective externalization in terms of supporting the generation of multiple hypotheses prior to making investigative

492 decisions, then cognitive short cuts such as satisficing and conformation bias might be
493 better managed.
494
495
496 **Key Points**

- We report the first empirical investigation of the use of decision logs by senior police detectives.
497 • The length and documentation style of decisions varied according to case type,
498 duration and the officer involved, some choosing minimal entries, some
499 making extensive entries. The analysis of logs indicates significant limitations
500 of the decision log format and guidance for supporting investigative decision-
501 making.
502 • Experienced SIOs generated more hypotheses early in the investigation and
503 switched between considering different hypotheses more often in the initial
504 and final phases of an investigation than inexperienced officers. Inexperienced
505 officers opened up more evidence sources than experienced officers early in
506 the investigation. These behaviors are consistent with higher levels of
507 satisficing and confirmation bias by inexperienced officers, and decreased
508 levels with experienced officers.
509 • The practical use of decision logs appears constrained by format and context
510 of use, arguably irreparably.

511 **Biographies**

512 **Coral J. Dando**

513 • Coral Dando is a Professor of Psychology at the University of Westminster,
514 London, a Forensic Psychologist, Chartered Psychologist and Chartered

Scientist. Her primary research interests are centered on applying psychological theory to understand and improve cognition in goal directed forensic settings. Coral was awarded her PhD in Psychology in 2008 by London South Bank University. Prior to commencing an academic career, Coral completed over 12 years service as a London police officer. She has written over 40 scientific journal articles, book chapters and commentaries, and her research has attracted approaching $2 million of funding from various bodies, including the UK and US governments.

Thomas C. Ormerod

- Thomas Ormerod is a Professor of Psychology and Head of School at the University of Sussex, UK. He is a fellow of the British Psychological Society, and was awarded his PhD in Cognitive Psychology in 1987 by the University of Sunderland, UK. Tom has studied expertise in naturalistic decision-making domains for over thirty years, and has published in excess of 100 peer reviewed scientific articles and book chapters. He has been principal investigator on research awards totaling in excess of $10m. His PhD research demonstrated about how computer-programming expertise can be understood in terms of theories of human reasoning.

References

ACPO. (2006). *Murder Investigation Manual.* Wybsoton: National Centre for Policing Excellence.

ACPO Crime Committee. (1999). Revised guidelines for the use of policy files. *https://www.app.college.police.uk/app-content/investigations/managing-investigations/* [Accessed 01 November, 2015]

Alby, F., Zucchermaglio, C., & Baruzzo, M. (2015). Diagnostic decision making in

oncology: creating shared knowledge and managing complexity. *Mind, Culture, and Activity*, *22*, 4-22.

Alison, L., Doran, B., Long, M. L., Power, N., & Humphrey, A. (2013). The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of Experimental Psychology: Applied*, *19*, 83.

Alison, L., van den Heuvel, C., Waring, S., Power, N., Long, A., O'Hara, T., & Crego, J. (2013). Immersive simulated learning environments for researching critical incidents: A knowledge synthesis of the literature and experiences of studying high-risk strategic decision making. *Journal of Cognitive Engineering and Decision Making*, *7*(3), 255-272.

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*, 43-63.

Ball, L. J., Maskill, L. & Ormerod, T. C. (1998). Satisficing in engineering design: Psychological determinants and implications for design support. *Automation In Construction, 7,* 213-22

Ball, L.J. & Ormerod, T.C. (1995). Structured and opportunistic processes in design: A critical discussion. *International Journal of Human-Computer Studies*, 43, 131-151.

Bichard, M. (2004). *The Bichard Inquiry report.* London: Her Majesty's Stationary Office.

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Medical Decision Making*, *35*, 539-557.

565 Chapman, G. B., Bergus, G. R., & Elstein, A. S. (1996). Order of information affects
566 clinical judgment. *Journal of Behavioral Decision Making*, *9*, 201-211.
567 Charmaz, K, & Henwood, K. (2007). Grounded theory. Chapter 14 in C.Willig & W.
568 Stainton-Rogers (Eds.). *The SAGE handbook of qualitative research in*
569 *psychology*. London: Sage.
570 Cohen-Hatton, S. R., Butler, P. C., & Honey, R. C. (2015). An investigation of
571 operational decision making in situ: Incident command in the UK Fire and
572 rescue service. *Human Factors*, *57*, 793-804.
573 Cohen-Hatton, S. R., & Honey, R. C. (2015). Goal-oriented training affects decision-
574 making processes in virtual and simulated fire and rescue environments.
575 *Journal of Experimental Psychology: Applied*, *21*, 395.
576 College of Policing. (2014). *Managing investigations.*
577 https://www.app.college.police.uk/app-content/investigations/managing-
578 Cox, R. (1999). Representation construction, externalised cognition and individual
579 differences. *Learning and Instruction*, *9*, 343-363.
580 Dougherty, D. M., Bjork, J. M., Moeller, F. G., Harper, R. A., Marsh, D. M., Mathias,
581 C. W., & Swann, A. C. (2003). Familial transmission of continuous
582 performance test behavior: attentional and impulsive response characteristics.
583 *The Journal of General Psychology*, *130*, 5-21.
584 Ellison, M., & Morgan, A. (2015). *Review of Miscarriages of Justice.* Home Office:
585 London.
586 Ericsson, K. A., & Simon, H. A. (1993) *Protocol Analysis: Verbal Reports as Data.*
587 MIT Press, Cambridge, MA.
588 Eyre, M., & Alison, L. (2008). To decide or not to decide: decision making and
589 decision avoidance in critical incidents. Chapter in D. Carson, R. Milne, F.

590 Pakes, K. Shalev, & A. Shawyer (Eds.), *Applying psychology to criminal*
591 *justice*, pp. 211-233.
592 Fahsing, I., & Ask, K. (2016). The making of an expert detective: the role of
593 experience in English and Norwegian police officers' investigative
594 decision-making, Psychology, Crime & Law, 22, 203-223. DOI:
595 10.1080/1068316X.2015.1077249
596 Home Affairs Select Committee. (2011). *Policing Large Scale Disorder: Lessons*
597 *from the disturbances of August 2011*. London: Home office.
598 Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
599 Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the
600 interrogation room: on the dangers of presuming guilt. *Law and human*
601 *behavior*, *27*, 187.
602 King, R.D., Whelan-K.E., Jones, F.M., Reiser, P.G.K., Bryant, C.H., Muggleton, N.H,
603 Kell, D.B., & Oliver, S.G. (2004). Functional genomic hypothesis generation
604 and experimentation by a robot scientist. *Nature, 427,* 247-252.
605 Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in
606 hypothesis testing. *Psychological Review*, *94*, 211- 228.
607 Leo, R. A. (2008). *Police interrogation and American justice*. Harvard University
608 Press: Harvard.
609 Magee, B. (2013). *Popper*. Routledge: Chicago.
610 Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic versus systematic processing
611 of information in detecting deception: Questioning the truth bias,
612 *Psychological Reports*, *105*, 11-36.
613 Meissner, C. A., & Kassin, S. M. (2002). " He's guilty!": investigator bias in
614 judgments of truth and deception. *Law and Human Behavior*, *26*, 469-480.

615 Mehle, T. (1982). Hypothesis generation in an automobile malfunction inference task.
616 *Acta Psychologica, 52*, 87-106.
617 Mossman, D. (2013). When forensic examiners disagree: Bias, or just inaccuracy?
618 *Psychology, Public Policy, and Law*, *19*, 40.
619 Ormerod, T. C., Barrett, E., & Taylor, P. J. (2008). Investigative sense-making in
620 criminal contexts. Chapter 5 in J-M Shraagen et al (Eds.), *Naturalistic*
621 *decision making and macrocognition*, Aldershot: Ashgate, pp. 81-102.
622 Rebitschek, F. G., Krems, J. F., & Jahn, G. (2015) Memory activation of multiple
623 hypotheses in sequential diagnostic reasoning, *Journal of Cognitive*
624 *Psychology*, 27, 780-796, DOI: 10.1080/20445911.2015.1026825
625 Runco, M. A. (1994). *Problem finding, problem solving, and creativity*. Greenwood
626 Publishing Group.
627 Sandham, A. (2013). *Hypothesis Generation in Investigative Scenarios*. Unpublished
628 PhD thesis, Lancaster University, UK.
629 Schulenberg, J. L. (2014). Systematic social observation of police decision-making:
630 the process, logistics, and challenges in a Canadian context. *Quality &*
631 *Quantity*, *48*, 297-315.
632 Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization
633 for situated reflection. *Cognitive Science*, *26*, 469-501.
634 Simon, H. A. (1956). Rational choice and the structure of the environment.
635 *Psychological Review*. *63*, 129–138. *doi*:*10.1037/h0042769*
636 Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*,
637 *41*, 1-20.
638 Steffensen, S. V. (2013). Human interactivity: problem-solving, solution-probing and
639 verbal patterns in the wild. In *Cognition Beyond the Brain*. London: Springer,

pp. 195-221.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, *2*, 492-494.

Tasca, M., Rodriguez, N., Spohn, C., & Koss, M. P. (2012). Police decision making in sexual assault cases: Predictors of suspect identification and arrest. *Journal of Interpersonal Violence*.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

van den Heuvel, C., Alison, L., & Power, N. (2014). Coping with uncertainty: police strategies for resilient decision-making and action implementation. *Cognition, Technology & Work*, *16*, 25-45.

Vickers, J. N., & Lewinski, W. (2012). Performing under pressure: Gaze control, decision making and shooting performance of elite and rookie police officers. *Human Movement Science*, *31*, 101-117.

Waring, S. K., Alison, L. J., Cunningham, S., & Whitfield, K. C. (2013). The impact of accountability on motivational goals and the quality of advice provided in crisis negotiations. *Psychology, Public Policy, and Law*, *19*, 137.

Wason, P. C. (1966). Reasoning. *New Horizons in Psychology, 1.* 135-151.

West Mercia Police (2015). *West Mercia police public protection arrangements discretionary serious case review Jamie Reynolds.* Shropshire Community NHS.

Wright, M. (2013). Homicide detectives' intuition. *Journal of Investigative Psychology and Offender Profiling, 10,* 182–199.