

**WestminsterResearch**

<http://www.westminster.ac.uk/westminsterresearch>

**Triage Decision-Making by Welfare Fraud Investigators**

**Walsh, D., Dando, C.J. and Ormerod, T.C.**

NOTICE: this is the authors' version of a work that was accepted for publication in Journal of Applied Research in Memory & Cognition. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in the Journal of Applied Research in Memory & Cognition, DOI: 10.1016/j.jarmac.2018.01.002, 2018.

The final definitive version in Journal of Applied Research in Memory & Cognition is available online at:

<https://dx.doi.org/10.1016/j.jarmac.2018.01.002>

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

---

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

---

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail [repository@westminster.ac.uk](mailto:repository@westminster.ac.uk)

Running header: Triage decision-making

Triage Decision-Making by Welfare Fraud Investigators

David Walsh,

Thomas C. Ormerod

Coral J. Dando

**Abstract**

Two studies explored triage decision-making in a welfare fraud investigation, specifically decisions concerning what evidence to collect when deciding whether to pursue a case to prosecution or whether to issue a warning only. An observational study revealed that triage decisions appear to be determined by subjective estimates of the ease of evidence collection, and that these estimates are influenced by complexity of mapping evidence onto fraud types. This hypothesis was explored in an experimental study of investigators, managers, and students choosing evidence to inform triage decisions for cases that varied according relevance and complexity. Students' selections were unaffected by the nature of the case. In contrast, with a simple fraud case, investigators and managers tended to select evidence to support a prosecution decision, but with a complex fraud case they selected evidence that supported comparative evaluation of prosecution and warning decisions. The results demonstrate flexible expertise in choosing what evidence to sample.

### **General Audience Summary**

Fraud is common in welfare claims, and costs taxpayers billions of dollars. Little is known about expertise in welfare fraud investigation, or more generally about the decision-making strategies used by crime investigators. The task of welfare fraud investigation begins with a triage decision: whether to seek evidence that might lead to prosecution or to terminate an investigation and issue a warning designed to dissuade clients from further claims. The first study explored the fraud investigation process, through field notes, interviews with investigators, and analysis of investigator interviews with clients. The study showed that investigators tended to make triage decisions, based on the perceived ease with which evidence might be collected. Decisions were affected by investigator experience and the ease with which they could understand how evidence would discriminate between cases.

A second experiment compared choices made by investigators, fraud case managers and students as to what evidence they would seek to make a triage decision. Participants were shown case scenarios based on a ‘tip-off’ that indicated evidence of fraud. Their task was to choose further evidence in order to decide whether to prosecute or issue a warning. The fraud cases differed in how easy it was to map evidence onto the underlying fraud activity. Students’ selections were unaffected by the nature of the case. In contrast, with a simple fraud case, investigators selected evidence to support a prosecution decision, but with a complex fraud case they selected evidence that could equally support prosecution and warning decisions.

The results suggest that investigators are flexible in the strategies they use in triage decision-making: they will seek evidence to discriminate between alternative case outcomes but only when the case is complex, otherwise they tend to seek evidence to confirm initial suspicions.

## Introduction

US Department of Labor statistics indicate fraud is found in 2.67% of welfare claims (Department of Labor, 2014). Welfare fraud costs the UK £1.2 billion per annum (Department for Work and Pensions, 2013), resulting in at least 350,000 investigations each year (Button, Shepherd, & Blackburn, 2016; Hansard, 2012; Walsh & Bull, 2013). Most fraud cases concern incorrect disclosure of personal circumstances (Walsh & Bull, 2013). For example, an unemployed individual might not disclose paid labor to avoid a reduction in benefit. Information that prompts the investigation of welfare fraud comes largely from anonymous 'tip-offs'. The quality of information in tip-offs ranges from highly detailed and substantively accurate to vague, inaccurate, and sometimes malicious.

There are approximately 5000 fraud investigators in the UK, working in teams of 25-30 in major cities (Smith, Button, Johnston, & Frimpong, 2011). An investigator is allocated each tip-off to investigate. Although procedural guidance exists, there is considerable discretion as to how the investigation should be managed, such as what information to gather, and when and how to close the investigation. Cases assessed as yielding robust evidence are selected for in-depth enquiries, to attempt to gain enough evidence for prosecution. Cases where it is deemed that evidence is unlikely to be gained are the subject of 'civil' rather than criminal investigations. Here, an investigator interviews the claimant to gain either an admission or denial. Admissions lead to a civil warning, with a termination of the benefit claim and repayment. Regardless of whether an admission or denial is obtained, this interview signals the end of a civil investigation. Investigators make their decisions based solely on details contained in the allegation, and their own estimation of the likely success of a prosecution. Choosing between criminal or civil

investigations is a key decision point because, once an investigator decides to resolve the case with a civil warning, switching to a criminal investigation is impossible because prosecution outcomes require that all evidence is collected with regard to relevant criminal legislation from the outset.

The number of fraud cases is large, and so a key task is triage - investigators must decide whether claimants should be prosecuted (a criminal investigation) or simply discouraged from making with a warning (a civil investigation). Triage involves decision-making under uncertainty, and is central to other domains such as medical diagnosis and criminal investigation. Little is known about the nature and flexibility of strategies used to make diagnostic decisions. Studies of expertise, ranging from chess (Simon & Gilmarin, 1973) to firefighting (Klein, 2004), have revealed experts often make rapid decisions based on recognition and retrieval of an action sequence. Novices, lacking relevant experience, rely instead upon general-purpose heuristics (Anderson, 2014), although experts may resort to general-purpose heuristics when a scenario is unfamiliar (e.g., DeStefano, Linstead, & Grey, 2011).

Here, we examine the strategies used by welfare fraud investigators to triage, and whether there are consistent biases in evidence selection. We report two empirical studies of welfare fraud investigation expertise: an observational study using interviews and work placement observations; and an experiment to test a hypothesis generated from the observational study concerning the relationship between experience and evidence prioritization.

### **Investigative Decision Making**

When individuals seek evidence to test a hypothesis, they should in principle gather diagnostic evidence that discriminates between competing hypotheses.

However, it is widely recognized that evidence selection shows confirmation bias, a

tendency to bolster the current hypothesis by seeking evidence that is consistent with it while disregarding inconsistent evidence (Nickerson, 1998). For example, Doherty, Mynatt, Tweney, and Schiavo (1979) found that, given one piece of evidence pertaining to a hypothesis, participants were more likely to seek new evidence about that hypothesis than to seek the same evidence for an alternative hypothesis.

If investigators favor an initial hypothesis to the exclusion of alternatives, this calls into question their ability to test hypotheses effectively. Even experienced investigators can make sub-optimal evidence selections. For example, criminal investigators (Meissner & Kassin, 2002) often assume guilt from the outset, referred to as a guilt bias (Kassin, Goldstein, & Savitsky, 2003). Even after extensive training designed to counter guilt bias, experienced investigators often still exhibit this behavior (Dando, Bull, Ormerod, & Sandham 2016). In insurance fraud investigation, the premature adoption of a hypothesis of guilt is a cause of investigative failure (Morley, Ball, & Ormerod, 2006). Similarly, welfare fraud investigators admitted to believing that clients with anomalous claims were guilty before interview, in contradiction of their training (Walsh & Bull, 2011). Where evidence is gathered that disconfirms guilt, it is judged as of lesser value and may be overlooked (Hasel, 2012).

Although evidence selection may be subject to bias, experienced investigators appear to develop expertise in hypothesis generation and testing. This has been demonstrated in police phone-call triage, scene-of-crime analysis, and hostage negotiation, where investigators adopt a range of strategies for generating and testing hypotheses depending on factors such as risk, criticality and frequency of the crime incident under investigation. (Dando & Ormerod, 2017; Fahsing & Ask, 2016; Ormerod, Barrett, & Taylor, 2008). Welfare fraud teams have a number of different roles. Case investigators follow leads, collect evidence, and interview the claimants.

Their managers make resource decisions (e.g., deciding what types of cases to focus on). As well as having more experience, all managers are promoted from an investigator role.

### **Study 1: Observational Study**

A mixed-methods approach was undertaken involving an ethnographic of benefits fraud investigation practices, and analyses of interviews between investigators and claimants suspected of fraud (a detailed description of these interviews is given by Walsh & Milne, 2008; Walsh & Bull, 2010). The aim of the study was to synthesize factors that investigators identify as key determinants of decision-making practice.

### **Method**

#### **Ethnography**

The first author conducted the ethnography. He was a senior investigations manager in the Department for Work and Pensions (DWP), the UK agency for social security benefits administration, involved in the investigation of social security benefit fraud for 20 years until 2006. The ethnography comprises both formal and informal field notes and interviews collected in the last five years of his role in DWP.

Interviews were conducted with 89 investigators ( $n=57$ ), investigation trainers ( $n=12$ ), national investigation inspectors ( $n=6$ ) and managers ( $n=14$ ). The overall mean years of experience of the investigators was 12.61,  $SD = 7.55$  (range 2-29 years' experience), trainers ( $M=17.83$ ,  $SD = 7.11$ , range 9-24 years), inspectors ( $M = 17.50$ ,  $SD = 5.79$ , range 8-23 years) and managers ( $M = 15.86$ ,  $SD = 5.13$ , range 7-25 years).



The interviews were semi-structured, and comprised conversations with trainers and managers, as well as performance appraisal interviews with 31 investigator line-managed by the researcher. The mean duration of the interviews was 50.66 minutes ( $SD = 18.10$ , ranging from 15-75 minutes). The interviews were analyzed, using thematic analysis (framed from an interpretive phenomenological basis), based on hand-written contemporaneous field notes. Questioning typically involved open enquiry designed to elicit reasons for undertaking a chosen course of action during investigations, based either on hypothetical scenarios or, in the case of investigators, actual cases. Examples of such questioning include:

- During an appraisal, researcher in his role as line manager, questioning an investigator: “What you would do given the claimant is reported to be working and claiming benefits and that work... he’s doing just occasional work for an unknown employer...where its.. is reported that ...when he does appear to go to work, he leaves between 7.30 am and 9.30 am”.
- During a case conference between investigators, in a case of a woman claiming benefits on the basis of her living alone whereas she was alleged to be living with her boyfriend: “So... why did you choose to interview the benefit claimant immediately rather than first gather documentary evidence such as utility bills or check other available records that might...might well have given some indication as to the length of time ... the two had been living together as a couple.. because that is providing retrospective evidence of fraud”.

### **Interviews with Claimants**

A sample of 241 recorded interviews conducted by investigators with suspects of welfare fraud between 1999-2006 were supplied, of which 20% (48) interviews were randomly selected for analysis. These interviews were coded and analyzed for

key themes using a Grounded Theory approach, which is well suited for this research because it supports the derivation of categories directly from the data rather than from pre-conceived hypotheses. The interviews were supplied directly to the research team following a request from the authors. The organization supplying the interviews were blind to the research aims and research questions, but had been provided with an overview document outlining the general nature of the research, how the data would be analyzed, stored, anonymized, and reported.

First, the audio data were transcribed verbatim, then the transcripts were passed to two coders who independently read each transcript. The focus then shifted to *open coding*, which involved coders working independently to identify behavioral concepts within the text, to saturation (Charmaz, 2006), and developing categories that represented their meaning in terms of properties and dimensions (Corbin & Strauss, 2008). Throughout, extensive notes were written to summarize the researchers' understanding, interpretations, and connections. Working together, using the research notes as a guide (Corbin & Strauss, 2008), categories created in 'open coding' were then refined to provide more precise and selective codes for behavior types. Throughout, behaviors were compared for similarities, violations, and differences within and between interviews. Although there were some initial disagreements between the three coders about particular categorizations, agreement was reached through a process of critical and constructive debate between coders, and the authors. Three primary codes emerged, which we have labeled i) evidence handling ii) account handling, and iii) legislation handling (see Table 1; Charmaz, 2006; Corbin & Strauss, 2008).

-----  
Insert Table 1 about here  
-----

### **Results and Discussion**

The observational study set out to establish the nature of fraud investigation practices. Table 1 summarizes nine key themes that emerge from the ethnography and interviews with claimants. The themes cluster into three categories: Evidence handling (collection, disclosure, exploration); Account handling (gathering, giving); and Legislation handling (explaining rights, obtaining information, establishing understanding, and inducing admissions). Each category is described as ‘handling’ because the categories show a range of strategic approaches adopted by investigators to deal most effectively with the evidence coming from a tip-off, the account given by (or to) a claimant, or the legal status of a claim.

Our observations revealed that case triage nearly always occurs in the very early stages of an investigation. For example, in one field note (identified under the ‘evidence gathering’ theme) the original allegation claimed that “he always has dirty hands when he comes into the office”. The investigator described how he inferred from this evidence the implication that the claimant was engaged in some form of manual work – such as mending cars, and that this “would be treated quite quickly as a case not worth pursuing”. In another case (identified under the ‘evidence exploration’ theme), where the tip-off suggested the claimant was working while claiming benefits, the investigation yielded a named employer who the investigator was familiar with. This led the investigator to state that the employer would keep records of employees’ earnings and would be co-operative, and that it was highly likely that the matter may be viewed as a potential prosecution case.

The examples above towards the poles of a continuum from vague to substantial suspicion. However, greater subjectivity in decision making was found where the originating allegations of wrongdoing lie towards the center. For example, a field note from an interview with an experienced investigator (coded under the ‘establishing understanding’ theme) suggested a key role for legislative practice in resolving uncertainty. The investigator stated that “those alleged to be working either as, say, a laborer for someone who is in turn self-employed would often reflect investigations that possessed no consistent pattern of evidence or information gathering, and that isn’t going to stand up in court, is it? In these ... cases I would either interview the benefit claimant immediately to seek their responses to the allegation, or if that failed undertake efforts to gather evidence or information; but mostly I’m going to discontinue the case early doors without any steps being taken... on the basis that the investigation would be futile”.

Even in cases where evidence had been gathered, investigators often made independent and quite different decisions as to how to conclude the case. For example, one field note (categorized under the ‘account giving’ theme) described a team meeting between a manager and three investigators working for her. Four different claimants had been under surveillance, which had identified them working as taxi drivers. Each case was assigned to a different investigator, but whereas two resulted in a prosecution, the others resulted in a civil outcome (a warning with termination of benefit claim; a requirement for repayment of wrongly paid benefit). This was despite these cases having the same initial characteristics. In each case, the investigators offered a hypothetical account of the circumstances of the claim to the manager as a way of justifying their decision.

The field notes reveal some inconsistency between investigators in responses to tip-offs that trigger an investigation, and also inconsistency of investigative activity once an investigation commences. For example, one field note (classified under the ‘account gathering’ theme) described a case where a welfare claimant was alleged to be working for a self-employed builder. The investigator originally assessed at the outset that he was potentially capable of gaining the evidence to prosecute. However, after failing to contact the builder employer, he decided the investigative activity necessary to gain wages records was not worth the effort, as the employer, even if they could be contacted, would “...likely be collusive, not keeping records but pay wages ‘cash in hand’ and is likely to employ him on an occasional basis anyway...”. The investigator stated in the interview with the researcher that they made this decision because in their judgment “...it is unlikely to yield the evidence needed for proof of criminal wrongdoing”. However, another investigator in a similar case to the above (possessing the very same characteristics) made a different judgment as to the case merits, continuing to investigate by undertaking surveillance, questioning the employer and advising him of his legal requirements in light of what was found after undertaking surveillance (in regard to how much work he was doing), or alternatively such surveillance did not reveal that he was working as alleged. Thus, the data show a level of individual differences leading to triage decision variability, with the consequent potential to spend time on cases not warranting such use of resources or to discontinue activity on cases that merit further criminal investigation still exists.

A number of investigators in field notes classified under the ‘evidence exploration’ theme described how, given that there is little guidance to assist them when making such case triage decisions, they use gut instinct to guide their decisions. They claimed their prior experiences enabled to make their decisions, citing a number

of reasons that influence these decisions, such as their estimated likelihood of gaining robust evidence to support a criminal prosecution. Examples involved cases such as a woman claiming benefits for her and her children who was alleged to be actually living with her husband or boyfriend. Some investigators painstakingly gathered documentary evidence to reinforce their suspicions.

Investigators often claimed, in field notes classified under the ‘obtaining information’ theme, to start with a ‘prosecution focused’ outlook, seeded by the description in the tip-off which necessarily emphasized the potential falsity of the claim and guilt of the claimant. Thus, making a ‘civil warning only’ decision was in effect over-riding a pre-existing ‘prosecute’ decision. However, despite such claims, it was evident from interactions between the researcher and the investigators that the point in the investigation when such decision making was made to abandon or continue towards a prosecution outcome was highly subjective and value laden.

The same variability was found when examining investigators’ reasoning for decisions. In a field note classified under the ‘establishing understanding’ theme, an investigator said that they discontinued with a prosecution focused case because surveillance was very difficult in that locality, for example it being impossible to observe in a covert manner, or that locals were very ‘surveillance aware’.

Alternatively, in a field note classified under the ‘evidence collection’ theme, documentary evidence obtained from a utility service provider was described as ‘insubstantial’ with the explanation that similar evidence had failed to secure a prosecution in prior investigations.

One of the managers’ tasks is to make decisions on the weekly ‘themes’ and to be alert to trends in claims: For example, one manager reported “We were seeing a lot of claim reports involving travelers, so we should drill down on this problem to try

and discourage it in our region”. This kind of decision-making reflects a kind of sampling and evaluation across cases, looking for commonalities of modus operandi, perpetrator and goal. The focus of investigators was almost exclusively on single current cases, the only exception being, as exemplified above, when reference was made to previous cases during evidence evaluation.

The way cases were dealt with depended on the perceived complexity of mapping evidence onto the suspected fraudulent activity. For example, in one field note (classified under the ‘account giving’ theme), a tip-off was evaluated by an investigator which suggested a claimant was receiving a welfare benefit through being deemed unfit for work, yet could be seen leaving their home each morning at the same time dressed in laboring clothes and carrying building tools. The investigator developed a narrative to explain the case: “So its straightforward, the tools show he’s working, the clothes, no-one’s going to choose them for social... like events`. The investigator went on to account how how ‘simple’ mappings between the evidence in the tip-off and the nature of the suspected fraud, would lead them to focus on collecting more evidence that might implicate the claimant: “With something as clear-cut as that, you want to spend your time looking beyond the tip-off for things that will corroborate your decision [to seek prosecution].”

Other tip-offs were less straightforward. For example, one field note (also classified under the ‘account giving’ theme) described a case in which the tip-off indicated that a woman was claiming welfare for children who were actually living away and looked after by her sister. The investigator stated “its likely she is not giving care because we know she has too much free time and parties too much to be looking after her kids”. The ‘evidence’ is not directly related to fraud and requires a belief-based inference to link the two. In cases such as this, investigators reported

resorting to more careful case reviews, and consultations with other colleagues. It was also suggested that the tougher the case, the more likely an investigator was to conclude that prosecution was unlikely and that a decision simply to issue a warning was more likely. One investigator reported “When you’ve got to think about it, it makes you think differently, and things become a little less obvious”. However, this phase of the study did not specifically measure for experience.

Experience was measured directly in examining the skill levels of investigators when they interviewed suspects. This analysis revealed that many of the less experienced investigators conducted interviews poorly (assessed on a five-point rating scale as below the median by the first author – see Walsh & Bull, 2010; Walsh & Milne, 2008 for details concerning the measurement scale, and its method of undertaking – along with details of measures of inter-rater agreement). One determinant of interview quality was whether sufficient evidence was gathered before the interview. It was often evident from examining those less skilled interviews that more evidence could have been gathered prior to the interview. Across the interviews a significant difference was found between the ratings for this task of preparation, where those investigators with more than 10 years’ experience ( $n = 108$ ,  $M = 2.77$   $SD = 0.84$ ) had higher ratings than those with less than this degree of experience ( $n = 133$ ,  $M = 2.43$ ,  $SD = 0.98$ );  $t(239) = 2.39$ ,  $p = .026$ ,  $d = 0.31$ .

An example of lower levels of skilled performance by less experienced investigators, classified under the ‘inducing admission’ theme, involved an interview of a woman by two male investigators who appeared to frighten the female suspect into admission of wrongdoing, despite presenting no evidence to her. Another field note classified under the ‘explaining rights’ theme involved a case where a less-experienced investigator relied on circumstantial evidence in the hope of an



admission, by suggesting to the claimant that if she withdrew her claim in the short term it would not affect her right to claim in the future (a wholly inaccurate statement as it transpired). It appeared as though the investigator was unwilling to exert the extra effort of gathering substantive evidence prior to the interview with the suspect, hoping that there would be full admission once circumstantial evidence was revealed.

In summary, the interviews and observations of Study 1 indicated that cases vary in complexity, specifically in the ease with which potential sources of evidence that are familiar from caseload experience can be mapped in a straightforward way onto the current case. Nine themes emerged that illustrate a range of strategies used to handle evidence, accounts of that evidence, and the legal framework in which evidence can be used. It is clear that an overriding issue is the extent to which evidence can be collected efficiently and used effectively, and estimates of this dominate the triage process. There was also some evidence of differing decision tactics according to the number of years of experience of investigators, and according to role (investigator versus manager). It is important to note that the only measure of 'expertise' per se that we were able to take was in analysis of the interviewer performance of investigators. Although this is suggestive of a correlation between years of experience and expertise, we cannot claim a definitive relationship. Nonetheless, the results of Study 1 suggest that both experience of the investigator and the apparent complexity of mapping evidence onto fraud type are important factors in triage decision-making. In particular, it appears that more experienced investigators make better use of evidence in interviews with claimants, and also distinguish between the utility of evidence types depending on ease of collection and verification.

## **Study 2: An Experimental Exploration of Triage Decision-Making**

In Study 2, we tested the hypothesised relationship between investigative experience, case complexity and triage decision-making in an experiment in which we presented participants with three scenarios: a simple suspected fraud case in which there was a direct mapping between the kinds of evidence presented to participants and the characteristics of the specific case under examination, and a complex case in which an inference was required to identify which piece of evidence might relate to a case. For each case, participants were given a description of the suspected fraud and a number showing the percentage of fraud cases where a specific piece of evidence was found that led to prosecution. As well as mimicking the real-life role of the “tip-off”, this piece of evidence played the role of ‘seed information’ that Doherty et al. (1999) used in their studies of hypothesis testing. The ‘seed information’ deliberately sets up a hypothesis that a case can be pursued to prosecution.

The participants’ task was then to select five further pieces of evidence based on the percentage of times the evidence led either to prosecution or to issuing of a warning only. A non-fraud case (choosing which of two streets a person might have left their car) was included to see if effects of experience in investigation were specific to cases relevant to the welfare fraud domain or generalized to other diagnostic decision-making tasks.

Within a single domain, it is possible to find different strategies and resulting biases, depending on the specific functions that individuals carry out. For example, Ormerod, Fritz, and Ridgway (1999) found that the categories into which expert textbook writers sorted tasks reflected superficial task characteristics (e.g., presence of graphics), whereas the categories into which expert designers of examinations sorted tasks reflected conceptual task characteristics (e.g., underlying

mathematical concepts). Our study examined two types of welfare fraud investigation experience: investigator and manager. This role distinction emerged from the ethnographic study, where it was found that managers undertook case comparisons tasks. Managers were also more experienced, and the interviews with claimants showed that more experienced investigation personnel were more skillful in conducting these interviews. We hypothesized that managers would pay more attention to alternative hypotheses than investigators, since their task involves comparing across cases rather than simply testing specific fraud hypotheses.

## Method

### Participants

A total of 146 participants took part, 73 investigators ( $M_{\text{investigative experience}} = 9.9$  years), 35 managers ( $M_{\text{investigative experience}} = 19.9$  years), and 38 domain-naive undergraduate and postgraduate students. The investigators comprised 43 females and 30 males ( $M_{\text{age}} = 43.81$ , ranging from 21 to 62 years), the managers 17 females and 18 males ( $M_{\text{age}} = 48.86$ , ranging from 21 to 68 years), and the students 16 females and 22 males ( $M_{\text{age}} = 31.89$ , ranging from 21 to 55 years). We chose to differentiate the manager and investigator groups by role rather than by years of experience because we could not assume a simple correlation between years of experience and expertise. Although the same is true for role, in every case the managers had been promoted after obtaining at least ten years investigation experience.

### Materials

The materials used were based on those developed by previous researchers (e.g., D'Addario & Macchi, 2012; Feeney et al., 2008; Mynatt, Doherty, & Dragan, 1993), and comprised a set of instructions and three diagnostic reasoning tasks presented in a six-page hardcopy booklet (see Appendix 1). The three problems in

each booklet were presented in one of six different orders. Each of the three problems was prefaced by a short summary of the background to the problem that described two potential case outcomes (equivalent to the alternative hypotheses under test in Mynatt et al. and others' studies). Start sentence that the case should be prosecuted, or that only a warning should be offered. Then one piece of base rate information was presented describing the percentage of instances of co-occurrence with a prosecute outcome. Thereafter, a further five pairs of categories of information were provided (see Appendix 3) in addition to the choice given in the preface to the problem. The instructions explained what was required of participants. They should select five pieces of information to help decide between alternative outcomes (to prosecute or issue a warning). On the next six pages of the booklet, each of the three problems was presented, comprised of a page showing the problem vignette and a page showing the evidence of co-occurrence choices. Finally, the booklet provided a debriefing statement.

Of the three tasks, two were relevant to the domain of fraud investigation, and one was domain-neutral (the 'house' problem from Feeney et al., 2008). The two fraud investigation tasks had a similar structure: A vignette was presented concerning a specific case and each case had six facts that were investigatively relevant.

- Task 1: a woman claiming benefits as a single parent who is suspected of living with and being supported financially by her boyfriend; (case details are that regular surveillance has showed the boyfriend leaving the claimant's house at the start of the working day and returning there at the end);
- Task 2: a man claiming benefits while working for a building company; (case details are that the claimant is thought to be working for a building company that hires people on a casual daily basis).

For each task, participants were required to select five additional pieces of evidence that would help them decide whether to follow an investigation strategy aimed towards prosecution.

The nature of a realistic triage task is that, while investigators might consider what evidence would be discriminatory, they do not actually access the evidence until an investigation is underway. Therefore, unlike other studies (e.g., Doherty et al., 1979), where participants were made aware of the percentage of co-occurrence after each selection was made, in the current study participants simply chose five items whose co-occurrence they would want to find out about.

The tasks differed in the degree of difficulty involved in mapping between case facts and co-occurrence evidence. In the simple task, the mapping between case facts and evidence choices was direct and required no inference. For case 1 above, the seed information was “the percentage of cases leading to prosecution in which surveillance has shown someone suspected of providing financial support to a single-parent claimant leaving and arriving at the address at times associated with the working day”. In the complex task, participants had to make inferences between the case facts and the evidence choices. For case 2 above, the seed information was “the percentage of cases leading to prosecution in which companies cannot produce detailed wage records for all their employees”. Such an evidence choice requires an inference to map onto the case fact that “companies that fail to keep wage records are also likely to be companies that hire on a casual daily basis”. The task was further complicated in that some pieces of evidence had a secondary inferential goal, requiring participants to identify whether the individual was likely to be a member of Gang A, in which a criminal investigation would be a likely outcome, or Gang B, in which a civil investigation would be the likely outcome. Complexity was varied in

this way to reflect the reality of cases faced by experienced investigators (Walsh, 2011). Each fraud task vignette was evaluated by four domain experts with > 20 years domain experience, who independently assessed the two case studies, using a five-point Likert scale for indices of case familiarity, realism and frequency. The vignettes achieved a score of at least 4 on each index by all four raters. Strong inter-agreement using Kappa measures was found across these scales (minimum 0.90).

### **Design and Procedure**

This research employed a mixed design, with Group as a between-subjects factor (investigators; managers; students) and Task (simple; complex; neutral) as a within-subjects factor. The experiment was presented during professional training days for the investigator and manager groups, and during lectures for the student groups (average group size on each testing occasion = 35 participants). Each participant first read the opening instructions on the problem booklet's front page (see Appendix 1 for task instructions). The order of task presentation was counter-balanced across participants. Furthermore, the order in which evidence items were listed was randomized. Participants were given five minutes to individually attempt each of the three tasks (which are presented in Appendix 2), making their selections using a pen, and without conferring with any other participant.

### **Results and discussion**

The data from 11 participants (4 managers, 3 investigators, and 4 students) were removed before analysis, because they failed to provide complete answers to one or more of the tasks.

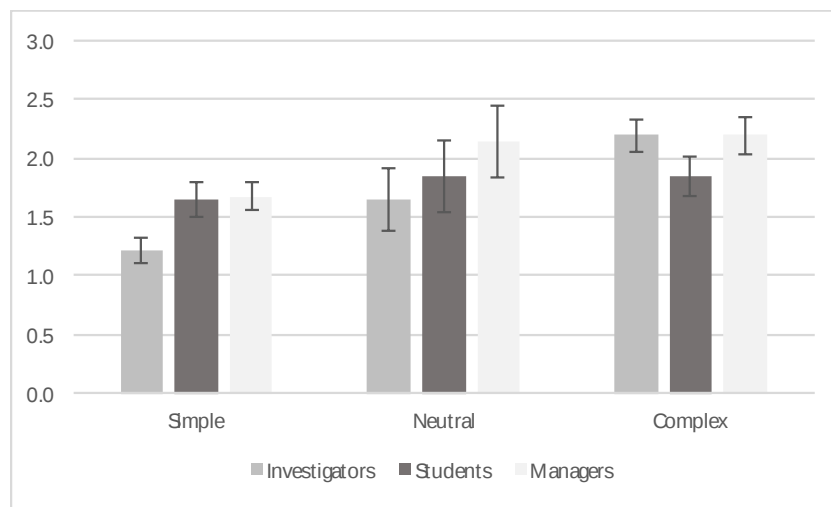
**Analysis of selections.** The answers provided by each participant were coded by the researchers according to the number of pairs of evidence selected. These were

then identified as follows: An evidence pair was chosen if the participant chose the same evidence item pertaining to each investigative outcome (i.e., the percentage of prosecutions given the specific evidence item AND the percentage of warnings only given the specific evidence item). Given that participants were given one evidence item in the initial vignette and then selected five more, the scores could vary between zero (no pairs – entirely selecting for the prosecute outcome seeded in the initial evidence) and three (all pairs – comparative across prosecute and warning outcomes for each piece of evidence they chose).

A Task (simple; complex; neutral) by Group (students; investigators; managers) ANOVA revealed a significant main effect of Task,  $F(2, 266) = 11.39, p < .001, \eta_p^2 = .08$ . Planned comparisons show that choices for complex tasks ( $M = 2.08$ ), 95% CI [1.92 2.24] showed significantly more pairs of prosecute/warning outcome selections than for simple tasks ( $M = 1.51$ ), 95% CI [1.36: 1.67],  $F(1, 135) = 30.62, p < 0.001, \eta_p^2 = .19$ . Choices for the neutral task ( $M = 1.88$ ) did not differ significantly from the other tasks,  $p = 0.59$ .

A significant main effect of Group was also found,  $F(1, 133) = 3.25, p = 0.04, \eta_p^2 = .05$  (see Fig. 1 for means and error). Managers, 95% CI [1.81: 2.23] and investigators, 95% CI [1.53: 1.84] choices did not differ significantly from those of students, 95% CI [1.56: 1.96],  $p = .46$ . However, managers' choices showed significantly more pairs of prosecute/warning outcome selections than those of investigators,  $F(2,133) = 6.38, p = .01$ .

Figure 1. Means for simple, neutral, and complex choices as a function of group.



A significant Task and Group interaction emerged,  $F(4, 266) = 2.54, p = .041$ ,  $\eta_p^2 = .04$ . Sheffé's post-hoc tests indicate that both managers and investigators chose significantly more pairs of prosecute/warning outcome selections for complex tasks than simple tasks,  $ps < .01$ , with no difference across tasks for students,  $p > .08$ .

**Analysis of selection consistency.** To examine whether the groups differed systematically in the types of evidence they selected, three analyses were conducted; one asked whether evidence selections made by participants within each Group were more similar than selections made across Groups, one examined whether Groups differed according to the types of evidence *content* they selected, and one compared the number of prosecution versus warning only (or King versus Queen Street) selections made for each of the tasks by the three Groups as a check for guilt bias (Kassin, 2012).

First, to examine whether the selections of participants reflected Group membership, we examined the mean distance of selections between pairs within groups (in-group mean distance) against between groups (out-group mean distance) across the three tasks. Mean distances were based on the squared Euclidian Distance



between pairs of participants as data and average linkage between groups calculated using hierarchical cluster analyses. To derive these means, we calculated the Euclidean distance for each participant pair, and then means for participant pairs within and between groups. Mean distances varied between 0 (absolute overlap in items selected) and 11 (no overlap at all in selections), and essentially indicate less or more variability in the selections made by participant pairs.

A Group (students; investigators; managers) by Task (simple; complex; neutral) by Proximity (in-group distance; out-group distance) ANOVA revealed a significant main effect of Task,  $F(2, 278) = 10.37, p < .01, \eta_p^2 = .07$ . Sheffé tests revealed that simple task selections ( $M = 4.75, 95\% \text{ CI } [4.65: 4.85]$ ) were more distant than complex task selections ( $M = 4.35, 95\% \text{ CI } [4.18: 4.52]$ ),  $p < .01$ . Neutral task selections ( $M = 4.54$ ) did not differ significantly from the other tasks,  $ps > 0.1$ . A significant main effect of Proximity was also found,  $F(1, 139) = 18.81, p < .011, \eta_p^2 = .12$ , with shorter distances within groups ( $M = 4.46, 95\% \text{ CI } [4.36: 4.57]$ ) than between groups ( $M = 4.63, 95\% \text{ CI } [4.53: 4.73]$ ). The interaction between Proximity and Group was also statistically significant,  $F(4, 266) = 2.54, p = .04, \eta_p^2 = .04$ . Sheffés tests indicate that student selections were significantly more distant from other groups ( $M = 4.73, 95\% \text{ CI } [4.60: 4.87]$ ) than within their own group ( $M = 4.43, 95\% \text{ CI } [4.29: 4.56]$ ),  $p < .01$ , with no difference in distances from own group and others for managers and investigators,  $p > .09$ . No other effects were significant,  $F_s < 2.15, ps > .09$ .

We then assessed whether the investigators and managers systematically selected response options different from those selected by the students. We summed the frequencies with which each evidence pair were selected (i.e., the ‘prosecute’ and ‘warning only’ pairs for each of the evidence contents in the fraud-related tasks, and

‘Queen Street’ and ‘King Street’ pairs in the neutral task). A series of Chi Square tests examined whether the groups differed in the frequencies with which they selected each evidence pair (five tests for each task, omitting the seeded evidence pair). Of these 15 tests, two were statistically significant (all others  $p > .12$ ). Investigators selected significantly more “percentage of houses in [King/Queen Street] with net curtains in the windows” evidence pairs in the neutral task than the other two Groups,  $\chi^2(2) = 6.45, p = 0.004$ . Similarly, investigators selected significantly more the “percentage of investigations, nationally that result in [prosecution/warning only], where the claimant had provided a false name for a suspected partner” evidence pairs in the simple fraud task than the other two groups,  $\chi^2(2) = 10.67, p = 0.005$ .

Third, to examine for evidence of guilt bias, we summed the number of ‘prosecution’ and ‘warning only’ pieces of evidence (or King Street and Queen Street for the neutral task) made by each of the three Groups. The number of ‘Prosecute’ and ‘King Street’ selections (i.e., the ones seeded in the initial evidence item presented to participants) made by each group for each task is shown in Table 2.

-----  
 Insert Table 2 here  
 -----

Overall, more seeded selections were made for the Simple Fraud task (64%) than for Complex Fraud task (51%) and the Neutral task (56%),  $\chi^2(2) = 14.12, p = 0.009$ . There were no differences between Groups for the any of the tasks, all  $ps > .12$ .

The results of Study 2 confirm that triage decision-making is affected by task complexity: when the task was simple, student participants tended to make evidence selections to test the ‘prosecute’ outcome seeded in the case description. When the

case was complex, the investigators and managers made more comparative selections than the students, choosing both prosecute and warning only outcomes for each evidence topic. As expected, students were unaffected by case complexity, and the three Groups did not differ significantly in their selections for the neutral case. These findings suggest that the effects are a result of domain experience rather than age, education or other demographic differences across groups.

### **General Discussion**

Two studies examined the possible nature of expertise in welfare fraud investigation. The first study used observational methods, and identified triage decision-making as a key aspect of the investigation role. Notably, it revealed how investigators make judgments about whether to pursue a case based on subjective feelings about which evidence for prosecution might be collected. The apparent simplicity of mapping the evidence of the tip-off onto kind of fraud suspected was reported as a key factor in decision-making. With simple cases, investigators reported a focus on seeking new corroborating evidence, while with complex cases, they reported a more evaluative and comparative approach, with more focus on considering alternative case outcomes. Any case is always open to alternative explanations. For example, while dressed for work with building tools is consistent with someone fraudulently working while claiming welfare benefits, it is possible, however unlikely, that the claimant was taking the tools each day to loan them to someone else. Thus, there is a risk of instinctive judgments in generating the lie bias found in other domains. The hypothesis generated in Study 1 concerning a relationship between experience of fraud investigation and case complexity was confirmed in the experiment conducted in Study 2.

In real fraud investigation situations, there are many sources of evidence that can be sampled in order to make a triage judgment. Just as DNA evidence is seen as more powerful than eye-witness evidence for achieving a conviction (e.g., Wells & Olson, 2003), fraud investigators may be aware that some pieces of evidence are more useful than others (e.g., bogus documentation is stronger evidence than an informant's testimonial). Thus, an alternative explanation for our findings is that the evidence selections made by investigators are determined by domain knowledge and context, that is, an understanding of what types of evidence are likely to prove most useful in securing a prosecution.

Some support for this hypothesis comes from an analysis of a mean Euclidean Distance measure of selections: Within the student Group selections were similar, whereas selections made by the expert groups differed. However, an analysis of the content of selections suggests that this Group similarity does not arise because experienced investigators were targeting specific evidence items. Only two out of 15 differences in evidence item selection were statistically significant. In one instance, investigators (when undertaking the simple task) selected more items than the other groups. That is, they more often chose "percentage of investigations, nationally, that result in [prosecution/warning only], where the claimant had provided a false name for a suspected partner". A putative explanation for this difference is that this is the only item where deceit (providing a false name) was explicit in the materials. If investigators are selecting to test a 'prosecute' hypothesis, then they should be influenced by the relevance of this evidence content. However, Manager participants did not all focus on this item.

The other significant difference was with the neutral task, where again investigators differed from the other groups in being more likely to select both the

“percentage of houses in [King/Queen Street] with net curtains in the windows” evidence items. Notwithstanding accounts of ‘nosey’ neighbors spying on welfare cheats through ‘twitching curtains’, it is stretching the data too far to interpret this difference as reflecting anything other than randomness. Overall, there is little evidence that selections were guided by knowledge concerning *known* evidence utility.

As Ormerod et al. (1999) noted, not all experts in a domain are the same kind of expert. Their study found differences between experienced designers of educational tasks and teachers of educational tasks, the latter tending to sort tasks according to conceptually deeper features than the former, despite the fact that task designers start off as teachers. In our study, across all three tasks, the managers’ selections were more similar than those of the investigators. Investigators’ and managers’ jobs differ in that the former are tasked with making decisions about individual cases (to prosecute or conduct a compliance interview only), while the latter are tasked with making resource prioritization decisions across a set of cases (i.e., they must tell investigators which cases from a current case load to focus their efforts upon). Thus, the estimated utility of evidence in a particular case will, for managers, be a judgment relative to the estimated utility of evidence in other cases. This additional layer of inferential complexity, we suggest, leads managers towards more comparative decision-making.

Our conclusions require certain caveats. We tested fraud investigation performance using only two fraud scenarios and two tasks. This limitation was imposed by the realities of working with domain experts, who cannot be taken off work tasks for extended periods to undertake research trials. Although the cases were pre-tested for familiarity, frequency and realism, it may be that small changes in case

content lead to large changes in diagnostic strategy. Further research may benefit from exploring a wider range of expertise domains and materials.

From a practical perspective, our data suggest that task complexity must be borne in mind when training investigators. One common approach used in police training is to attempt to ‘de-bias’ individuals by alerting them towards the negative consequences of confirmation bias. A similar de-biasing approach was advocated by Kern and Doherty (1982) in the domain of medical diagnosis. In our data, we saw some evidence for a guilt bias, in the preference for selections by all groups for the ‘prosecute’ evidence in the Simple Fraud task. However, this bias disappeared with the Complex tasks. These data indicate that de-biasing through instruction may be not be necessary, whether or not it is appropriate for investigative training, since any ‘bias’ is highly case-specific and potentially normative for the specific context in which it arises. Instead, training should focus upon identifying the right hypothesis-testing strategy for each case.

### References

- Anderson, J. R. (2014). *Rules of the mind*, (2<sup>nd</sup> Edition). Hove: Psychology Press.
- Button, M., Shepherd, D., & Blackbourn, D. (2016). *The Fraud ‘Justice’ systems; A scoping study on the civil, regulatory, and private paths to ‘justice’ for fraudsters*. Main report: Portsmouth: University of Portsmouth
- D’Addario, M., & Macchi, L. (2012). Pseudodiagnosticity: the role of the rarity factor in the perception of the informativeness of data. *Psychology*, 3, 489.
- Dando, C. J., Bull, R., Ormerod, T. C., and Sandham, A. (2013). Helping to sort the liars from the truth-tellers: The gradual revelation of information during investigative interviews. *Legal & Criminological Psychology*, 20, 114-128. doi:

10.111/lrcp.12016

Department for Work and Pensions (2013). *Annual Report and Accounts 2012-13*.

London: The Stationery Office.

Department of Labor (2014). Improper payment data for the Unemployment Insurance (UI) program derived from the Benefit Accuracy Measurement (BAM) program for the 12-month period ending March 31, 2014.

<http://www.dol.gov/dol/maps/Data.htm>. Retrieved on 22 January 2015.

Destefano, M., Lindstedt, J. K., & Gray, W. D. (2011). Use of complementary actions decreases with expertise. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *33rd Annual Conference of the Cognitive Science Society* (pp. 2709-2014). Austin, TX: Cognitive Science Society.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979).

Pseudodiagnosticity. *Acta Psychologica*, *43*, 111–121. doi:10.1016/0001-6918(79)90017-9

Fahsing, I., & Ask, K. (2013). Decision making and decisional tipping points in homicide investigations: An interview study of British and Norwegian detectives. *Journal of Investigative Psychology and Offender Profiling*, *10*, 155-165. DOI: 10.1002/jip.1384

Fahsing, I., & Ask, K. (2016). The making of an expert detective: The role of experience in English and Norwegian police officers' investigative decision-making. *Psychology, Crime & Law*, *22*, 203-223. doi: 10.1080/1068316X.2015.1077249

Feeney, A., Evans, J., & Venn, S. (2008). Rarity, pseudodiagnosticity and Bayesian reasoning. *Thinking & Reasoning*, *14*, 209–230.

doi:10.1080/13546780801934549

Hansard (2012). Available at

[http://www.publications.parliament.uk/pa/cm201213/cmhansrd/cm121015/text/121015w0001.htm#121015w0001.htm\\_wqn24](http://www.publications.parliament.uk/pa/cm201213/cmhansrd/cm121015/text/121015w0001.htm#121015w0001.htm_wqn24) (last accessed 28th January 2015).

Hasel, L. E. (2012). Evidentiary independence. In L Nadel & W. Sinnott-Armstrong (Eds.), *Memory and Law*, (pp. 142-158). Oxford: Oxford University Press.

Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist*, 67, 431-45. doi: 10.1037/a0028212.

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: on the dangers of presuming guilt. *Law and Human Behavior*, 27, 187-203.

Kern, L., & Doherty, M. E. (1982). 'Pseudodiagnosticity' in an idealized medical problem-solving environment. *Academic Medicine*, 57, 100-4.

Klein, G. (2004). *The power of intuition*. New York: A Currency Book/Doubleday.

Loveland, I. (1989). The micro-politics of welfare rights: The interpretation and application of the cohabitation rule in the Housing Benefit scheme. *Journal of Social Welfare Law*, 11, 23-42.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": investigator bias in judgments of truth and deception. *Law and Human Behavior*, 26, 469-480.

Morley, N. J., Ball, L. J., & Ormerod, T.C. (2006). How the detection of insurance fraud succeeds and fails. *Psychology, Crime and Law*, 12, 163-180.

Moore, P. (1981). Scroungermania at the DHSS. *New Society*, January 22.

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *Quarterly Journal of Experimental Psychology Section A*, 46, 759-778.



doi:10.1080/14640749308401038

- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175-220. doi: [10.1037/1089-2680.2.2.175](https://doi.org/10.1037/1089-2680.2.2.175)
- Ormerod, T. C., Barrett, E. & Taylor, P. (2008). Investigative sense-making in criminal contexts. In J-M Shraagen, L. Militello, T. Ormerod, & R. Lipshitz (Eds.) *Macrocognition and Naturalistic Decision Making* (pp. 81-102). Aldershot, UK: Ashgate.
- Ormerod, T. C., Fritz, C. O. & Ridgway, J. (1999). From deep to superficial categorisation with increasing expertise. *Proceedings from the 21<sup>st</sup> conference of the Cognitive Science Society*, Vancouver, August 1999, pp.502-506.
- Simon, H. A., & Gilmarin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, 5, 29-46.
- Smith, R. (1985). Who's fiddling: Fraud and abuse. In S. Ward (Ed). *DHSS in Crisis* (pp. 130-146). London: CPAG.
- Smith, G., Button, M., Johnston, L., & Frimpong, K. (2011) *Studying fraud as white-collar crime*. Basingstoke: Palgrave Macmillan.
- Walsh, D. & Bull, R. (2010). The interviewing of suspects by non-police agencies: What's effective? What is effective! *Legal and Criminological Psychology*, 15, 305-321.
- Walsh, D. & Milne, R. (2008). Keeping the P.E.A.C.E.? An Analysis of the taped interview performance of benefit fraud investigators within the DWP. *Legal and Criminological Psychology*.13, 39-57.
- Walsh, D. (2011). *Toward a framework for the interviewing suspects of fraud*. Unpublished PhD thesis: University of Leicester.

Walsh, D., & Bull, R. (2013). The investigation and investigative interviewing of benefit fraud suspects in the UK: Historical and contemporary perspectives. In B. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 33-58). New York: Springer.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277-295.

**Table 1. Themes from a grounded theory analysis of study 1 field notes**

<b>Evidence Handling (Collection; Disclosure; Exploration)</b>
<p><b>Evidence collection</b> - Substantial documentary records collected, some of which either may not necessarily have supported suspicions, but sufficient documentary evidence (such as utility bills, loan agreements, tax records) had been collected that in their totality strongly suggested that the woman had been living with her boyfriend for a number of years.</p> <p><b>Evidence disclosure</b> - Clear indications of a plans undertaken by the investigator prior to the interview to introduce the evidence in a gradual manner with, for example, the less strong evidence presented first followed incrementally by increasing strong evidence as the interview progressed with the records that showed that the woman had lived with her boyfriend for a number of years presented towards the end to eventually counter initial responses such as <i>"OK – I do admit that he has moved in with me but it only has just occurred"</i></p> <p><b>Evidence exploration</b> - No (or insufficient) evidence presented in the interview and the result largely depended on the co-operation of the suspect admitting to the alleged fraud. A typical question would be <i>"We have had a report that you have been working for the past few months helping a local builder – what have you to say about that allegation"</i></p>
<b>Account Handling (Gathering; Giving)</b>
<p><b>Account gathering</b> - Interviewer at first gathers an account from the suspect that seeks their version of events in response to the original allegation and then displays active listening (as the interview develops) that takes into account those initial responses. A typical exchange in the interview (in a case of a woman suspected of living with her boyfriend) would be <i>" Bearing in mind you said earlier that you had never lived with him, can you explain why last year when he has renewed his car insurance that he declared he was living at your address?"</i></p> <p><b>Account giving</b> - Interviewer does not seek an account but provides their own view or suggests a scenario reflecting their own opinion. For example, the interviewer would typically say in the early stages of the interview <i>"I have loads of evidence that proves to me that you have been living together for at least 6 months, so you might as well admit it now"</i> In such cases the interviewer did not seek initial explanations or listen to any refutations but proceeded to even suggest why the suspect had not declared that the boyfriend had moved in'</p>
<b>Legislation Handling ( Explaining rights; Obtaining information; Establishing understanding; Inducing admissions)</b>
<p><b>Explaining rights</b> - Ensures that the suspect understand their rights in the interview before it begins. <i>"It is important I know that you understand the rights that I have just read out to you before I continue. As such, can you explain your understanding of them. Don't worry if you get them wrong, they can sound quite complicated. I know I had trouble understanding the when I first heard them"</i></p> <p><b>Evidence gathered under correct protocols/legislation</b> - <i>"Ok I have used (cites correct legislation) which authorizes me to collect that information for the purposes of investigation of crime, so I am afraid you are mistaken to suggest that your privacy has been breached by my asking your boyfriend's employer for details of whether his wages are paid into a joint bank account, address, next of kin etc."</i></p> <p><b>Establishing understanding</b> - <i>" OK so I need firstly to ensure that you understand your knowledge of the benefits system, why did you claim benefits in the first place?"</i></p> <p><b>Inducing admissions</b> - <i>"it's much better for you to tell me now that you have done wrong"</i></p>

Table 2.

*Frequency (% in brackets) of selections consistent with the presented hypothesis*

*(‘prosecution’/ ‘King Street’) by each group for each task*

	Neutral Task	Simple Fraud Task	Complex Fraud Task
Students	156 (53)	178 (62)	131 (52)
Investigators	118 (64)	123 (69)	87 (50)
Managers	77 (50)	95 (61)	79 (53)

**Appendix 1: General task instructions**

Thank you for agreeing to take part in this study. You will be asked to complete a number of tasks that require you to think about the evidence you need to gather in order to choose between two alternative hypotheses. The study should take around 15 minutes in total. Your participation in this study is voluntary, and you have the right to withdraw at any stage. The data will be held anonymously. We are interested in how decisions vary according to the type of task undertaken. We will be recording the responses you select for each task. We want you to focus on the tasks without interruption, but you should not rush since the choices you make are more important to us than the speed with which you make them. In this booklet you will find three tasks. Each task requires you to select some sources of evidence that test a hypothesis. Some of these tasks are about *benefits fraud*, where individuals deliberately claim a benefit, such as invalidity or unemployment, to which they know they are not entitled. All of the names in the tasks are fictitious.

Imagine you work for the Department of Work and Pensions (DWP) as an investigator. Your job is to examine potential cases of fraud. Specifically, in each case, you must determine whether it is worth collecting evidence to seek prosecution of the suspected fraudster, or whether you will simply interview the individual to encourage them to stop claiming benefits without taking them to court. An investigator needs to judge whether the likelihood of securing a prosecution is worth the time and effort invested in collecting evidence, because collecting evidence can cost a lot of taxpayers' money. However, a prosecution can sometimes be the only way of preventing repeated offending. So, you will need to choose pieces of evidence that might help you decide whether a case is likely to turn out to lead to prosecution, or whether a case should simply be completed by interview.

## Appendix 2: Task materials

You can now request five of the following eleven pieces of information to assist you in determining whether to pursue an investigation with a view to prosecution or claim termination (but no more):

- 1) % of investigations that result in a prosecution, where credit reference checks show that a male has stated that his home address is that of the benefit claimant.
- 2) % of investigations where, in a previous investigation the benefit claimant provided a false name for a suspected male, and a further investigation leads to prosecution.
- 3) % of cases which result in claim termination (but no more), where credit reference checks show that the benefit claimant and a male have both stated (on a joint application for credit) that they live at the same address.
- 4) % of investigations, where the employer holds a different address for the male, that result in a claim termination (but no more).
- 5) % of investigations where the owner of the vehicle is identified as belonging to the suspected boyfriend that result in prosecution.
- 6) % of investigations where, in a previous investigation the benefit claimant provided a false name for a suspected male, and a further investigation results in claim termination (but no more).
- 7) % of cases resulting in claim termination (but no more), where surveillance established that a male regularly leaves the property of the benefit claimant.
- 8) % of investigations that result in a claim termination (but no more) , where credit reference checks show that a male has stated that his home address is that of the benefit claimant.
- 9) % of investigations where, regardless of what address the male's employer holds, the cases result in prosecution.
- 10) % of cases which result in prosecution, where credit reference checks show that the benefit claimant and a male have both stated (on a joint application for credit) that they live at the same address.
- 11) % of investigations where the owner of the vehicle is identified as belonging to the suspected boyfriend that result in claim termination (but no more).

Task 2

Non-fraud case

Before you went to the pub last night you parked your car. You know that you left it in one of two streets, either Queen St or King St, but it is now the next morning and you can't remember which.

You remember the following details of the house outside which you left the car last night:

The house had uPVC windows

The house had a front garden

The house had a two-car driveway

The house had a security light

The house had a "beware of the dog" sign outside

The house had net curtains in the windows

The council database shows that 70% of the houses in Queen St have security lights.

You can now request five of the following eleven further pieces of information to assist you in determining whether you left your car in Queen St or King St:

- 1) The % of houses in King St with net curtains in the windows
- 2) The % of houses in Queen St with uPVC windows
- 3) The % of houses in King St with a front garden
- 4) The % of houses in Queen St with "beware of the dog" signs outside
- 5) The % of houses in King St with uPVC windows
- 6) The % of houses in Queen St with a two-car driveway
- 7) The % of houses in King St with security lights
- 8) The % of houses in Queen St with a front garden
- 9) The % of houses in King St with "beware of the dog" signs outside
- 10) The % of houses in Queen St with net curtains in the windows
- 11) The % of houses in King St with a two-car driveway

Task 3

Complex fraud case

You are trying to identify which of two gangs working on two different sites a benefit claimant is with. It is alleged that he has been working with the gang at various sites continually for almost a year. One of the gangs (Gang A) work for a building firm that is collusive and unreliable and several of the gang are thought to be claiming benefits , while the other gang (Gang B) are working for a building firm that are very co-operative and keep excellent wage records, although it is possible that some members of that gang are also claiming benefits.. You know that the benefit claimant lives within a mile of where the work is being done but you do not know exactly where the work is being done. Whichever gang is involved, they are both undertaking the plastering. The benefit claimant has stated, when signing as unemployed, that he is a qualified painter and decorator and that is the line of work he is looking for. Two outcomes only are possible. If he has been working for Gang B, you will undertake an investigation to bring about a prosecution. On the other hand, if he is working for Gang A then it is likely that you will look to investigate the case with a view to obtaining a claim withdrawal from the claimant (but no more)

Please choose five of the pieces of information, by circling the relevant numbers.

You have the following pieces of information:

65% of members of Gang A live within 1 mile of the building site.

Eleven additional pieces of information are also available:

- 1). % of members of Gang A who are claiming benefits.
- 2). % of members of Gang A who are decorators.
- 3). % of claimants who work for reliable employers where investigations result in a claim withdrawal (but no more).
- 4) % of claimants who work for unreliable employers who are prosecuted after an investigation.
- 5) % of cases resulting in claim withdrawal (but no more) that involve claimants working as painters and decorators.
- 6) % of members of Gang B who live within 1 mile of the building site.
- 7) % of members in Gang B who are claiming benefits.



- 8) % of claimants who work for reliable employers who are prosecuted after an investigation.
- 9) % of members of Gang B who are decorators.
- 10) . % of claimants who work for unreliable employers where investigations result in a claim withdrawal (but no more).
- 11) % of cases resulting in prosecution that involve claimants working as painters and decorators,

**Appendix 3: Percentage of selections of each item of information for each task made by each group (S = students, I = Investigators, M = Managers).**

1. Non-fraud task

Evidence (% of houses in...)	% overall	S n=66	I n= 41	M n=35
King St with net curtains in the windows	33	20	15	10
Queen St with uPVC windows	30	19	16	6
King St with a front garden	27	20	9	7
Queen St with “beware of the dog” sign	83	52	32	28
King St with uPVC windows	13	8	5	5
Queen St with a two-car driveway	74	45	27	28
King St with security lights	61	37	22	23
Queen St with a front garden	33	18	19	8
King St with “beware of the dog” sign	72	49	21	27
Queen St with net curtains in the windows	39	22	24	7
King St with a two-car driveway	61	40	16	27

## 2. Simple fraud task

Evidence (% of investigations, nationally that...)	% overall	S n=66	I n= 41	M n=35
result in prosecution, where another person gives claimant's address.	59	36	27	17
result in prosecution, where claimant provided a false name for a partner.	55	28	30	16
end in interview, where claimant and partner have joint credit application.	50	36	14	18
end in interview, where employer holds a different address for a partner.	24	10	9	13
result in prosecution, where another person owns vehicle at the claimant's address.	57	35	21	21
end in interview, where claimant provided a false name for a partner.	28	14	18	6
end in interview, where surveillance shows partner leaves claimant's house each day.	64	40	27	19
end in interview, where surveillance shows partner leaves claimant's house each day.	36	24	10	14
result in prosecution, where employer holds a different address for a partner.	42	28	18	11
result in prosecution, where claimant and partner have joint credit application.	80	51	27	30
end in interview, where another person owns a vehicle at the claimant's address.	28	25	4	9

## 3. Complex fraud task

Evidence (% of)	% overall	S n=66	I n= 41	M n=35
cases leading to prosecution in which companies cannot produce wage records.	72	44	28	25
members of Gang A who are painters and decorators.	64	40	24	23
members of Gang B claiming welfare benefits.	21	12	10	6
cases leading to prosecution where companies fail to meet regulatory standards.	30	21	12	7
cases leading to interview where companies employ casual labourers out of normal hours.	22	10	15	5
members of Gang B live within one mile of the building site.	83	54	31	27
cases leading to interview only in which companies cannot produce wage records.	70	42	26	27
members of Gang A claiming welfare benefits	24	12	12	9
members of Gang B who are painters and decorators.	70	44	23	27
cases leading to interview only where companies fail to meet regulatory standards.	24	13	14	6
cases leading to prosecution where companies employ casual labourers out of normal hours	30	14	11	15