

# IF THEY GET IT RIGHT: AN EXPERIMENTAL TEST OF THE EFFECTS OF THE APPOINTMENT AND REPORTS OF UK PUBLIC INQUIRIES

---

RAANAN SULITZEANU-KENAN

A number of arguments regarding the politics of UK public inquiries (PIs) suggest that the appointment of a public inquiry and its subsequent report affect public responsibility attributions in ways that could be beneficial to the appointing office holder. One claim refers to the effect of an appointment on responsibility attribution towards the appointer of a PI; another refers to the relative strength of the effects of PI reports on responsibility attributions compared with other public evaluations. This latter argument relies on the assumption that PIs are judged as more credible than other conveyors of public evaluations. To test these hypotheses, this research employs two web-based experiments involving a sample of 474 UK citizens. The findings do not support the hypotheses. Instead, they reveal that the credibility of PIs is conditional upon acceptability of the report content.

## INTRODUCTION

Following a policy failure which raises public concern, will the appointment of a public inquiry reduce the responsibility attributed to the appointing government minister? In terms of credibility, does a public inquiry report enjoy more credibility than political speeches? Does it affect public opinion more than textually equivalent political speeches? Given that so many public inquiries are labelled as 'whitewash' by the general public, how is it that these *ad hoc* institutions are sought by the public, almost reflexively, in the aftermath of adverse events? This study presents empirical answers to the first three questions and relies on the findings and on the case of the Hutton Inquiry, appointed in July 2003 into the circumstances surrounding the death of Dr David Kelly, a scientist employed as an advisor by the Ministry of Defence, to suggest an answer to the fourth.

Public inquiries may come into existence in the aftermath of negative events such as disasters, accidents, policy failures and scandals. In the UK, the decision of whether to have a public inquiry (PI) in the wake of such an event typically lies in the hands either of individual ministers or the government collectively. A number of explanations account for the employment of these *ad hoc* institutions. These explanations rely on certain empirical assumptions regarding public perceptions of these bodies and public reactions to their outputs – in particular, the credibility attributed to them by the public and the relative influence of their reports on public opinion. However,

Raanan Sulitzeanu-Kenan is a DPhil candidate in politics at Wolfson College, Oxford University.

none of these assumptions have yet been tested. This study presents the results of such experimental testing.

The term 'public inquiry' (PI) is a loose one (Public Administration Select Committee (PASC) 2005, p. 7), often used to denote different types of institutions and functions. Important distinctions can be made, for example, between planning, advising and investigating PIs. The following criteria will be used to define PIs in the context of this study:

1. An *ad hoc* institution: that is, one established for a particular task; once its primary task is concluded, the tribunal is dissolved;
2. Formally external to the executive;
3. Established by the government or a minister;
4. As a result of the appointer's discretion: that is, not the result of a requirement prescribed by any statute or other rule;
5. For the main task of investigation: a criterion used to distinguish between investigative and advisory functions (Weare 1955, pp. 43–4);
6. Of past event(s);
7. In a public way: that is, it is not only directed *inward* (to the appointing body) but also *outward*, to the public, typically during a crisis of confidence between the public and government (Wade and Forsyth 1994, p. 1007), in a way which allows exposure of relevant facts to public scrutiny (Clarke 2000, p. 8).

PIs that qualify for the above criteria come into existence in a variety of circumstances and they range from fairly routine investigations of specific types of misfortunes to unique and extreme disasters as well as fiascoes. Much like the recent UK Public Administration Select Committee's (PASC) Report (2005, p. 7), this study focuses on those PIs set up to investigate specific controversial events that have given rise to public concern. However, more selectively, this study concentrates on the particular cases in which the office holder, with the power to decide whether to set up a PI or not, is themselves included in the group of those potentially responsible for the negative event to be investigated. Such a condition excludes from the analysis PIs into 'historical events', that is, events which occurred under a previous government unless some implications of the affair – for example, participation in an ongoing cover-up – raises concern regarding the responsibility of current office holders. Examples of such cases are the second inquiry into 'Bloody Sunday' (set up in 1998 by the Blair Government into the events on Sunday 30 January 1972 during the Conservative Heath Government) or the inquiry into the Marchioness disaster (set up in 2000 by the New Labour Blair Government into the events of August 1989 during the Conservative Thatcher Government). However, more selectively, this study concentrates on the particular cases in which the office holders, with the power to decide whether to set up a PI or not, are themselves included in the group of those potentially responsible for the negative event to be investigated. The remaining set of cases includes PIs that conform to the above criteria and are not 'historical', for example, the 1936

Porter Inquiry into unauthorized disclosure of information relating to the Budget and the 1996 Scott Inquiry into sales of arms to Iraq. In recent decades, such inquiries were typically chaired by serving or retired judges (65.5 per cent of the notable inquiries since 1990: PASC 2005, p. 19), though also by retired senior civil servants and prominent lawyers.

## THE APPOINTMENT OF A PUBLIC INQUIRY

### Restoring confidence and learning in the aftermath of a crisis

The employment of an independent tribunal of inquiry, formally external to the executive and legislative institutions, in the aftermath of crisis, disaster and scandals, has been called a 'British contribution to the legal world' (Segal 1984, p. 206). Such an inquiry is handed the task of thoroughly investigating the affair in question, and providing a public account of it. Such *ad hoc* institutions first appeared in British public life in the late nineteenth century (Keeton 1960, pp. 64–5), and acquired formal status in the 1921 Tribunals of Inquiry Act (which provides the legal basis for the most powerful form of inquiry).

Between the years 1921 and 2005 only 24 such inquiries have been established. Although the Act provides a statutory procedure for establishing public inquiries, it coexists with other specific (statutory) inquiries (for example, under the Railway Act 1974; the Merchant Shipping Act 1970; the Police Act 1964). Furthermore, according to legal convention, the 1921 Act did not exclude non-statutory inquiries, those established under Crown prerogative by government ministers (for example, the 1983 Franks Inquiry into the UK Falklands Policy, and the 1997–2000 Phillips Inquiry into the BSE affair).

The basic argument underlying these inquiries is that crises, disasters and scandals often result in public disquiet and in loss of confidence in the body politic. Confidence can be effectively restored only by thoroughly investigating and establishing the truth and exposing the facts to public scrutiny (Salmon 1966; Clarke 2000, p. 8).

Another related argument focuses on the function of learning as an important aspect of PI activity. Learning is a common reason given for the appointment of PIs, that is, learning for the purpose of drawing lessons which would help prevent or reduce the chance of a similar event happening (Howe 1999; Clarke 2000, p. 7; Elliot and McGuinness 2002). The relative advantage of PIs in learning and establishing thorough accounts of complex affairs have been noted by various observers (*The Times* 13.12.1988; Polidano 1999).

### Credibility issues

As the argument goes, a PI is thought to be effective in restoring confidence only if it enjoys the public's trust (McLean 2001, p. 592). Only a *credible* body of investigation will do. Indeed, the enactment of the 1921 Act has been viewed as resulting from profound dissatisfaction with the adequacy and credibility of political forums for investigating matters of great public concern and which

have caused embarrassment to the government (Salmon 1966; Segal 1984; Woodhouse 1995; Howe 1999). The 1966 Salmon Commission on Tribunals of Inquiry thus concluded that select committees were not suitable for 'the investigation of allegations of public misconduct... Such matters should be entirely removed from political influences' (Salmon 1966). The inadequacy of parliamentary committees for these investigations stems from the dominance of the executive in the British legislative system (Drewry 1996; Flinders 2001).

Much of the literature about the appropriate form public inquiries should take concerns requisites to safeguard their credibility. Salmon's second recommendation in his 1966 report states that members of the PI should be of the highest standing; in addition, he recommends that the chairperson should hold a high judicial office while other members should have no connections with any political party (Salmon 1966). Others stress only the importance of externality of PI members from government (Polidano 2001) or from any political party (Segal 1984).

### **The shortcomings of public inquiries (PIs)**

Despite these requirements, Elliot and McGuinness have pointed to an 'ongoing failure to learn' by PIs (2002, p. 16). Their critique is divided into three aspects: impartiality, process and underlying purpose. By impartiality, they refer to three further aspects. The first implies a broader interpretation of the term impartiality to include aspects other than overt partisan affiliations or direct interest in the subject of the inquiry. Elliot and McGuinness address the fact that judges are not representative of the entire population, but rather of a narrow segment of society. The typical social background and professional education of judges emphasize particular values and this leads them to act in broadly similar ways when faced with such political situations – as 'protectors and conservators of what has been, of the relationships and interest on which, *in their view* [sic], our society is founded' (Griffith 1991, p. 57). The second shortcoming of impartiality appears to stem from political influence at various stages of the work of PIs in order to affect the outcomes of an inquiry (Woodhouse 1995, p. 26; McLean and Johnes 2000a, b). As might be expected, direct indications of this influence are limited, and often obtained only many years after the fact. The third aspect of impartiality suggests that 'framing' also plays a role in the work of PIs. Framing is defined as selecting 'some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation' (Entman 1993, p. 52). One example of this, given by Elliot and McGuinness, is the role of the term 'hooliganism' in a series of football disaster inquiries – something which distracted the discourse from issues of 'crowd safety' (2002, p. 17).

It could be argued that the functions of credible investigation for restoring confidence as well as for learning in order to reduce the likelihood of reoccurrence, can be seen as normative functions of PIs, that is, as the appropriate

outcomes of a successful PI. However, given the fact that PIs are creations of politicians, it is not surprising that a number of writers suggest a descriptive function of PI – and one that is at odds with, and indeed an impediment to, attaining the two normative functions – the employment of PIs for the purpose of blame avoidance.

### **Anticipatory and reactive blame avoidance**

Weaver's (1986) argument that blame avoidance is a dominant political motivation of policy-makers rests on negativity bias in electoral behaviour. Negativity bias refers to the findings which suggest that losers are more likely to notice and act on their grievance than gainers are to act on the basis of their improved state (Bloom and Price 1975; Kernell 1977; Lau 1985). Weaver suggests that, under such conditions, policy-makers 'are not credit-claiming maximizers, but blame minimizers, and credit-claiming and "good policy" satisficers' (1986, p. 372). Apart from negativity bias, another political danger of 'politicized failures' lies in the opportunities they pose for opponents to call attention to problems and to provide critics of policy-makers with a signal that 'something is rotten' (Kingdon 1995, p. 98; Brändström and Kuipers 2003, p. 281). These works give general support to the hypothesis that blame-avoidance motivation would play a dominant role in policy-maker's choices, particularly when blame has been introduced into the public sphere and it is directed 'close to home', as if often the case in situations that give rise to PIs.

Kathleen McGraw (1991) has commented on Weaver's (1986) work that 'avoidance of a blame generating situation is not always possible', and politicians may find themselves in a situation for which they will be blamed. McGraw refers to these 'after-the-fact damage control tactics as blame management strategies' (1991, p. 1135). Similarly, in Hart's review of Richard Ellis's *Presidential Lightning Rods* (1994), he notes the failure to clearly distinguish between keeping out of trouble and deflecting blame for trouble not avoided (1995, p. 1022). Arnold (1990) explicitly makes this distinction while noting that US Congress members may recognize the potential criticism certain policies can bring about. Such policies are voted against whether or not concrete opposition exists, 'in *anticipation* of future punishment, not in response to current pressures' (1990, p. 75; italics added). It appears that we should distinguish between two distinct phases of blame-avoidance activity.

In the first, the political act occurs prior to either the emergence of blame or the occurrence of a blame-generating event, that is, as an anticipatory act. In the second, the political act comes after the emergence of blame, that is, as a reactive act. The two phases of blame-avoidance activity differ in their temporal arrangement with respect to the emergence of blame. It is suggested here that this distinction is employed in the analysis, that is, that the term 'anticipatory blame avoidance' (ABA) is used for political strategies which are employed prior to the generation of blame, and that the term

'reactive blame avoidance' (RBA) is used to denote strategies employed after the introduction of blame (Sulitzeanu-Kenan and Hood 2005).

### **Agenda control and responsibility judgements**

When shifting the focus of analysis from strategies to their expected effects, a clear distinction can be drawn between strategies that attempt to directly affect the public's responsibility attribution judgements, and strategies that endeavour to influence the public agenda. This distinction is similar to the difference between influencing 'what people think' (responsibility judgement), and 'what people think about' (agenda control) (Entman 1989). When one attempts to empirically study these effects, this distinction is particularly relevant from a methodological aspect. Agenda effects can be measured by media coverage, parliamentary discussions, and legislative proposals. Responsibility attribution judgements can be gauged through public opinion questionnaires and interviews (Sulitzeanu-Kenan and Hood 2005). The latter techniques are also much more amenable to controlled experimental designs. Indeed, the hypotheses to be tested in this research are those that predict responsibility attribution effects.

### **Public inquiries (PIs) as a reactive blame-avoidance strategy**

There are a number of reasons to suggest that judicial inquiries may be a mechanism employed by ministers to evade and deflect responsibility. A starting point for a discussion on the role of PI in blame avoidance should be that the decision to establish a PI in the first place rests in the hands of individual ministers or in the government collectively (Flinders 2001, p. 163). Indeed, the literature addressing the politics of public inquiries in the UK, as well as in other countries, suggests that a PI may hold some blame-avoidance advantages for the appointing government or office holder. These advantages can be divided into short-term advantages – following the appointment – and long-term advantages – resulting from PI reports.

In the short term, an appointment may: (1) remove the issue from the political arena and reduce the sense of urgency by creating an image of de-politicization, and an impression of decisive action (Lipsky and Olson 1977, pp. 443–4; Woodhouse 1995, p. 25; Bovens *et al.* 1999); (2) 'buy time' for the government, allowing public anger to abate, elections to take place, ministers to be either reshuffled, promoted to the House of Lords or persuaded to leave politics (Flinders 2001; Brändström and Kuipers 2003); (3) give the appearance that ministers share public concern: that is, that they could not have been personally involved. The appointment implicitly frames the government as 'not included' in the potential 'wrong-doers group' (Woodhouse 1995, p. 25); and (4) 'block' alternative investigative procedures (for example, of parliamentary committees or criminal investigations) – as a result of rules and conventions governing conflicts of institutional investigative authority and freedom of speech and particularly of the press (for

example, *subjudice*) (Kremnitzer 1989; Flinders 2001, p. 164; Elliot and McGuinness 2001, p. 21; Polidano 2001).

Although all of these claims may be categorized as reactive blame-avoidance activities attained by appointing a PI, a closer look points to three different operating logics. The first two claims refer to agenda-control advantages of removing or reducing the intensity of debate over the inquired event, thereby gaining time to address the problem. The third, suggested by Woodhouse (1995, p. 25), implies an effect of the appointment on public opinion, that is, that the appointment of a PI reduces the level of responsibility attributed to the appointing office holder. The fourth advantage is an institutional one, resulting from legal rules and traditions about the relative hierarchy and relationships between various institutions of inquiry, and its rationale is dependent on the relative risk posed by the different institutions of inquiry.

The experiments employed in this research are designed to measure opinions. They therefore support the testing of hypotheses that predict effects on responsibility attribution judgements rather than agenda control. For this reason, the claim to be tested regarding the short-term effect of PI appointment is the one suggested by Woodhouse (1995).

*Hypothesis 1:* The immediate appointment of a public inquiry once an affair becomes public will result in lower responsibility attribution to the appointing minister, compared with the same situation without an appointment.

In the long term, blame-avoidance advantages rely on greater acceptance of PI evaluations (reports) of an affair compared with evaluations made by other public figures such as politicians. This argument, advanced by Drewry (1975, 1996), depicts the employment of PIs as 'borrowed authority'. The notion of 'borrowed authority' was first introduced by Drewry (1975) under the phrase 'harnessing a myth', and later became one of the most common explanations for the political motivation to employ judicial inquiries.

The political advantage lies in judges' *capacity ex officio to legitimate decisions in a manner denied to other people*. They enjoy a head start over rival contenders by wearing a mantle of almost unquestionable authority, enhanced by an image of political neutrality, which both them and those around them have been socialized into taking more or less for granted. That is the nub of politicians' case for capitalizing upon the myth that judges are set apart from politics. (Drewry 1975, p. 56; italics added)

The 'borrowed authority' argument rests upon the following elements: (1) judges have a unique public image of independence and hence, credibility; (2) this image enables them to retrospectively 'legitimate' official action (Drewry 1975, p. 58).

As in the first 'normative' function of PIs – investigation for restoring confidence – the 'borrowed authority' thesis also rests on the credibility condition. Yet unlike the 'normative' function, this thesis purports to be a 'descriptive' one – delineating the harnessing of judicial credibility in attaining the effect of mitigating responsibility attribution.

The claim regarding judges' unique public image is the main explanation for the choice of senior judges to chair major inquiries into political scandals, major public calamities and crises of regime-legitimacy. Other explanations refer to judges' professional expertise (Department of Constitutional Affairs 2004, pp. 21–2), yet most writers see the main issue as being the image of judges as independent, impartial, unbiased and, thus, credible (Drewry 1975, 1996; Woodhouse 1995, p. 25; Thompson 1997, p. 183; Flinders 2001, p. 161; Elliot and McGuinness 2002, p. 17; Woolf 2004).

A more sceptical view of the 'independence and impartiality' argument was expressed in a 2005 Public Administration Select Committee (PASC) Report on public inquiries. The committee attributed this to the growth of judicial review in recent years, something which increased the involvement of judges in politically controversial cases, and to media interest in judges' ideological positions. The committee inferred from the reactions to the Hutton Inquiry that:

Those who do not agree with an inquiry's conclusions may not perceive it as independent and objective, regardless of whether the chair is a member of the judiciary. . . . If their [judges] reports fail to conclude that minister and senior officials are to blame, they may be heralded as a 'whitewash' by political opponents and the media . . . . (PASC 2005, p. 22)

The PASC Report, however, refers only to elite reactions (politicians and the press) and leaves the question of the general public's perception of PI reports open.

Drewry leaves some vagueness regarding the second element of the notion of 'borrowed authority'. How does the appointment of a trusted public figure (typically a judge) help 'defuse matters'? One understanding of this ability to diffuse matters is in the context of responsibility attribution, that is, that the enhanced trust enjoyed by judges allows their public evaluations to carry more weight on public opinion. Such a reading of Drewry is supported by his claim that '(i)t is almost a reflex action at times of dire political emergency for judges to be rushed to the scene to extinguish the blaze (*often by retrospectively legitimating official action*) and spread calm and reassurance' (Drewry 1975, p. 58; italics added).

However, if one accepts this mechanism – which links trust with influence on public attribution of responsibility – that the greater authority, impartiality and credibility enjoyed by judges may also be detrimental to the interests of government. If the PI's report about the inquired affair is a positive evaluation of the government, this evaluation carries more weight since it was made by a trusted figure. But if the report is negative, it could be more



harmful than other (non-judicial) evaluations of the government. Roy Hattersley (2000) expresses this idea clearly:

Politicians can abuse each other without serious people taking very much notice. But if a high court judge, master of an Oxbridge college or retired senior civil servant exposed the processes of government as rotten, our comfortable world would never be the same again. (Hattersley 2000)

Thus, the second claim to be tested in this experiment is an elaboration of the 'borrowed authority' thesis. Such a claim relies on two arguments: (1) That an evaluation made by a public inquiry enjoys a higher level of credibility; (2) Therefore, its content has a greater effect than other political evaluations on responsibility attribution. In more formal terms, this hypothesis predicts a particular relationship between two factors and responsibility attribution. The first is an institutional factor – which varies between the condition in which the conveyer of evaluation is a PI, and the condition in which it is a political figure; the second is an evaluation content factor – which varies between an exonerating evaluation (positive) condition, and a critical evaluation (negative) condition. The first part of the hypothesis predicts that credibility will mediate the effect of the institutional factor on responsibility attribution. In its second part, the hypothesis predicts a moderation model of the effect of the two factors on responsibility attribution (Baron and Kenny 1986; Glenberg 1996, p. 400; Shadish *et al.* 2001, p. 264), that is, that the effect of evaluation content on responsibility attribution will vary across different conditions of the institutional factor, as shown in figure 1.

*Hypothesis 2:*

- (a) PI reports enjoy higher credibility than non-PI evaluations (that is, an institutional factor main effect on credibility judgement);
- (b) The effects of PI reports on responsibility attribution are greater than the effects of non-PI evaluations (that is, an interaction effect of institutional factor and evaluation content factor on responsibility attribution).

## RESEARCH DESIGN

The method chosen to test these hypotheses is experimental. A combination of the two reasons posits an experimental design as the best method, both in light of the nature of the phenomenon (low 'signal-to-noise' ratio effects) and under data availability restrictions. An experimental design allows us to selectively assess the effects of particular factors on the dependent variables, with a high level of internal validity, by controlling for many other 'noise' factors. For example, public acceptance of a PI report may also be influenced by media acceptance and the responses of other elite figures (PASC 2005). An experiment can control for these factors, providing a highly

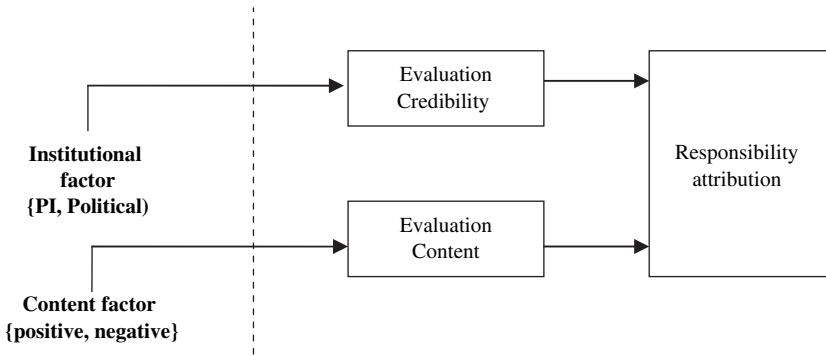


FIGURE 1 A graphic description of Hypothesis 2

effective method of studying the above hypotheses. Second, other research methods, for example, medium-N analyses of real-life PI cases, are seriously restricted for researching these hypotheses due to lack of systematic data (public opinion polls) on the dependent variables – public responsibility attribution and credibility judgements.

The research involves two experiments which rely upon an experimental vignette. The vignette includes a short report, the length of and in the style of a newspaper article, about a health regulation failure. This story was identical for all experimental groups; however, the story was accompanied by additional information and this varied between groups. In the additional information, one dichotomous factor was manipulated for the first experiment, and two dichotomous factors were manipulated for the second experiment. Table 1 reports the attributes that were varied in the stories.

The two experiments relied upon six web-based surveys (two for Experiment 1, and four for Experiment 2), administered to a sample of 474 UK citizens. Respondents were drawn from the pool of registered respondents of SurveySchack.com Ltd, and were randomly assigned to the six experimental groups (for sample characteristics, see appendix 1). After their assignment to a particular group, each respondent received an email invitation to answer that group's designated questionnaire. The email invitation included a coded link, which led the respondent to the assigned questionnaire. Respondents were unaware that there was more than one version of the questionnaire and, after completing it, a respondent was technically restricted from answering the questionnaire again. It should be noted that web-based surveys, though ill-suited to estimate population characteristics (means and proportions), do allow valid inferences about relationships, especially when the most important variables of interest are based on controlled experimental treatments that are randomly applied to respondents (Berrens *et al.* 2003, p. 2). This also allows the prevalence of the use of 'convenience samples' for such purposes.

TABLE 1 *The structure of vignette's experimental manipulations*

<b>EXPERIMENT 1</b>		
<b>Manipulation</b>	<b>Versions</b>	<b>Hypotheses</b>
Immediate appointment	A. No mention of appointment	Lower responsibility attribution to the minister when a PI has been appointed
	B. Reporting on the appointment of a PI right after the failure story	
<b>EXPERIMENT 2</b>		
Institutional factor (PI or 'political' evaluations)	A. The failure story is followed by a summary of a PI report	An interaction between the institutional factor and evaluation content factor: The effect of evaluation content on responsibility attribution is larger under PI evaluation than under 'political' evaluation
	B. The failure story is followed by a summary of a political speech	
Evaluation content factor (negative or positive evaluations)	A. A positive evaluation of the Department of Health's handling of the affair	
	B. A negative evaluation of the Department of Health's handling of the affair	

### Enhancing external validity

Unlike many experiments of this type, this study did not rely on undergraduate students but rather on a varied sample of the British public. However, although the sampling method used satisfies the requirement of random assignment, it fails to satisfy the requirement of random sampling. Furthermore, the sample does not constitute a representative sample of the British public (see also appendix 1).

In order to enhance the ability to generalize from the findings of the experiment, a 'sample re-weighting' method was used (Shadish 2002, p. 386). Respondents were asked to reply to a set of demographic questions. Analysis of the results indicated that both gender and voting intentions were significantly related to responsibility attribution. For this reason, all experimental groups were re-weighted for gender and voting intentions in order to have these proportions equal in all experimental groups and similar to the proportions of these characteristics in the general UK population. (It should be noted, however, that all significant effects found in the study were found for both un-weighted and re-weighted samples.)

All the groups were re-weighted to approximately 50:50 proportions of women and men. For voting intentions, the question used in the experiment to measure voting intentions was borrowed from the Mori Political Attitudes Poll ('How would you vote if there were a General Election tomorrow?'). The choices included 'Conservative', 'Labour', 'Liberal-Democrat' (LD), 'Other', and 'Would not vote'. The closest Mori Political Attitudes Poll to the time of the experiment was the one taken on 4 August 2004 (12 days after the beginning of the experiment and 5 days after its completion), and its results were used to adjust the experiment sample in order to approximate the following proportions: Conservative – 26 per cent, Labour – 30 per cent, LD – 17 per cent, Other – 9 per cent, and 'Would not vote' – 18 per cent. The resultant re-weighted sample (N = 522) is more representative of the British population and was the one used for the analysis.

### **Selecting the experimental 'failure story'**

A number of considerations shaped the process of selecting the failure story in the attempt to create a valid simulation. Validity in this context means that it was necessary that the event be a feasible one, and one which might occur in British public life; in addition, it would be an event that would be perceived as a possible candidate for PI investigation. For this purpose, the failure story was selected from the general set of such events which took place in recent decades. A further, somewhat contradictory, consideration addressed the need to avoid a 'familiar' story, which could influence the judgement of respondents, for example, an affair which was consequently investigated and reported upon, and could cause 'contamination by information' (Fischhoff 1975; Fischhoff and Beyth 1975) to respondents who remembered it.

The first step was a survey of a sample of the national British press between 1983 and 1993. Candidate cases were defined as those deemed serious enough for a call to be made for the appointment of a 'public (or) judicial inquiry', and which had passed the 'gate-keeping' mechanism of editorial decisions in at least one of the sampled newspapers. The most recent ten-year period was deliberately avoided in order to exclude events that were fresher and therefore more memorable. The search resulted in a large number of events and these provided a pool of valid events for simulation.

The second selection consideration led to favouring an event that would not be 'too' distinct or unique to be remembered – and narrowed the options to public failure of the 'departmental error' type (using the categorization of Dowding and Kang 1998). From this list of events, the event selected was the withdrawal of the anti-arthritis drug Oprelvekin, which prompted calls for a public inquiry in July 1985. (It should be noted here that the experiment was conducted between the 23rd and 30th of July 2004, well before the recent Vioxx anti-arthritis drug affair, which was first publicized in the UK on 1 October 2004.) The affair followed the withdrawal of Oprelvekin in 1982, after the deaths of 83 people and almost 4000 cases of side-effects. The Oprelvekin

'action committee' claimed that the affair was a comprehensive failure by the Department of Health and the Committee on Safety of Medicines (CSM) – the government's drug safety watchdog. About 1000 claimants were expected to become involved in claims for damages against the CSM, the Department of Health and the manufacturers. One opposition MP, in a letter to Prime Minister Thatcher, demanded a public inquiry. (The newspaper article on which the failure story is based is given in appendix 2.)

Since the experimental event was to be used to test responsibility attribution effects, in particular to ministers, Dowding and Kang's (1998) work was consulted to estimate the ministerial responsibility associated with such an affair. Dowding and Kang found that departmental errors are associated with 15 per cent of 'resignation-issues' (defined as events which led to calls for the resignation of a minister that were reported in the national press), and with 5.6 per cent of actual resignations. The proportion of actual resignations from departmental-error which became resignation-issues is 13 per cent (1998, p. 421). Additionally, Dowding and Kang found that 24 per cent of the events involving 'personal error' ended in resignation, and that 50 per cent of 'financial scandals' resulted in resignations. Relying on these findings, and in order to increase the likelihood of responsibility attribution to the minister, the story was then modified. First, a statement by the minister in support of the initial approval of the drug was added in order to create the impression that he was aware of the approval decision. Secondly, the adverse financial consequences of the affair were elaborated upon in order to include elements of a financial problem in addition to the core affair. These considerations led to the construction of the experimental vignette as a departmental error, which involved an indication of personal knowledge on the part of the minister, and included financial aspects (for the 'failure story' used in the experiment, see appendix 3).

### *Experiment 1*

Hypothesis 1 was tested by a simple experimental design. Respondents (N = 237) were randomly assigned to one of two groups. Both groups 1 and 2 were presented with the failure story.

*Manipulation of the appointment factor:* respondents in group 2 were also informed that the minister had appointed a PI, in the following form:

The Minister of Health yesterday told the House of Commons that he has appointed a public inquiry, headed by Lord Justice Fields, a senior Court of Appeal judge, to investigate all aspects of the affair.

The wording was intended also to control for the implications of the 'terms of reference' of PIs. Indeed, the terms of reference appear to have an effect both on the acceptance and impact of an appointment (Brändström and Kuipers 2003, p. 297), and of the PI's report. In the Hutton Inquiry, the inquiry's terms of reference and their interpretation by Lord Hutton were a major part of the public debate that followed the publication of the report.

By choosing the widest terms of reference in the experiment, the design avoids any such complications. This entails that the experiment cannot test the effect of various terms of reference on public opinion; nor was it intended to do so. Any significant difference between experimental groups would be reasonably interpreted as resulting from the appointment of a PI with widely defined terms of reference.

### *Experiment 2*

Respondents (N = 240) were randomly assigned to one of four different groups representing a 2 (*institutional factor*: PI evaluation versus political evaluation) × 2 (*evaluation content factor*: positive vs. negative) between-subjects factorial design. All four groups were presented with the failure story.

*Manipulation of the institutional factor*: next, respondents read a paragraph that reported on a public evaluation of the affair. In the PI evaluation condition, participants read:

The Minister of Health has appointed a public inquiry, headed by Lord Justice Fields, a senior Court of Appeal judge, to investigate all aspects of the affair. The conclusion of the inquiry report was announced yesterday by Lord Fields:

In the political evaluation condition, participants read:

In a debate in the House of Commons yesterday the Minister of Health stated his position on the affair:

or:

In a debate in the House of Commons yesterday the Leader of the Opposition stated his position on the affair:

for the positive evaluation, and the negative evaluation, respectively (table 2).

It should be noted that insofar as other professionals, such as retired senior civil servants or distinguished lawyers, may enjoy an image of political independence and credibility, Hypothesis 2 may apply to PIs chaired by them as well. However, since it is assumed that judges still represent clear cases of politically independent public figures (and indeed are referred to as such in the literature), and in order to facilitate the clearest possible inferences

TABLE 2 *Manipulation arrangement in Experiment 2*

		<b>Institutional factor</b>	
		<b>PI report</b>	<b>Political speech</b>
<b>Evaluation content factor</b>	<i>Positive</i>	A positive PI report	A positive evaluation in a political speech
	<i>Negative</i>	A negative PI report	A negative evaluation is a political speech

from the findings, the experiment utilizes a judicial inquiry for testing these hypotheses.

*Manipulation of the evaluation content factor:* following the text which set the institutional context of the evaluation, respondents were presented with two different evaluations of the affair, one positive – exonerating the Department of Health, and the other negative – critical of the department. In the positive evaluation condition, participants read:

In 2001, when Nerpo was approved in the UK, as well as in 11 other European countries, there was no scientific information available which indicated a risk in using the drug. Therefore, the decision to approve the use of Nerpo was reasonable at the time. Two years later, after the American warning about the use of Nerpo, the Agency for Safety of Medicines within the Department needed some time to independently evaluate the risk for patients in the UK in order to make a responsible decision about withdrawing the drug in this country. Some risk is always involved in approving new medicines, but the overall policy of handling the safety of medicines by the Department of Health is responsible, and properly managed.

In the negative evaluation condition, participants read:

The 2001 decision by the Department of Health to approve the use of Nerpo was a mistake. Although the scientific information available now was not known at that time, a more prudent approach would have been to wait for more information before approving the drug. Two years later, after the American warning about the use of Nerpo was published, the Agency for Safety of Medicines within the Department had failed to take action for nearly a whole year. This delay prolonged the use of Nerpo and possibly caused additional damage, which could have been avoided. This series of mistakes suggests that a more general problem exists in the ability of the Department of Health to adequately evaluate and react to new information about public health risks.

## Dependent variables

### *Responsibility attribution*

The term 'responsibility' holds a wide range of meanings. H.L.A. Hart (1968) distinguished between four senses of 'responsibility': (1) role; (2) causal; (3) liability; and (4) capacity. In the context of psychological research, where an actor's responsibility for an outcome is questioned, some researchers equate responsibility with the ascription of cause, whereas others equate it with assignment of sanctions (Hagiwara 1992, p. 145). Indeed, blame attribution is strongly correlated with punishment-worthiness (Gibson and Gouws 1999, p. 512). The concept of blame-attribution is often used in the literature to denote responsibility attribution and vice versa. Rudolf (2003) has advised avoiding the words 'blame' or 'credit' in measurements of responsibility

attributions. In this paper, the term responsibility attribution is the one that is consistently used.

A further categorization has been suggested by Iyengar (1989) to distinguish between two dimensions of responsibility, namely causation and treatment. Agents identified as agents of causation (of adverse situations) are negatively evaluated, and those identified as agents of treatment are positively evaluated (Iyengar 1989, p. 895). The types of measurement scales used in responsibility attribution research include both closed-ended (Iyengar 1989; Gibson and Gouws 1999; Rudolf 2003; Arceneaux 2003), and open-ended scales (Haider-Markel and Joslyn 2001, p. 530). For reasons of measurement reliability between experimental groups, the experiment employs a closed-ended scale.

The measurement of responsibility attribution was performed with respect to three hierarchical levels by asking separately about the responsibility of the minister, the Department of Health (or its director general) and the agency for the safety of medicines (or its head). Three items were used to measure responsibility attribution, by asking separately about responsibility (minister/department/agency), need for resignation (minister/director-general/agency-head), and need for reform (department/agency). The hierarchical dimension was used to assess various distributions of responsibility in the different scenarios tested (for example, the expected relative reduction of ministerial responsibility attribution following the appointment of a PI). Apart from direct questions about responsibility attribution, the questions about 'need-to-resign' were an adaptation of 'punishment-worthiness', and the questions about 'need-to-reform' were an adaptation of the 'treatment responsibility' concept.

### **Credibility and accuracy judgements**

As noted above, the credibility of evaluations was expected to be an important factor of Hypothesis 2. For this purpose, respondents in Experiment 2 were asked whether the evaluation (PI report or political speech) was biased, using a ten-point scale. The use of 'bias' was preferred to a direct question about credibility; this was based on comprehension problems with the latter question encountered in pilot work. Furthermore, the notion of 'credibility' in question is that of 'neutrality' and 'independence' (as opposed to 'competence'). The complementing values of this measure were used as the credibility scale. This item was expected to show a main effect of institutional factor, that is, that PI reports would be judged as more credible than political speeches.

Another item was used to measure respondents' judgement about the accuracy of the evaluations, as a representation of the affair, using a ten-point scale. This item was aimed at assessing the acceptability of evaluation content more independently from its institutional source. Such an assessment is particularly important since the experiment relies on one specific 'failure story'. The accuracy judgement would indicate the level of acceptability of the two evaluations, given the mean reaction to the failure story.



### **Manipulation check**

In order to assess the internal validity of experimental designs that make use of vignettes, it is necessary to determine whether the manipulations were successful at generating different perceptions among respondents who heard alternative versions of the vignette (McGraw 1996; Gibson and Gouws 1999, p. 506). It was assumed that respondents correctly understood the report on the appointment of a PI (Experiment 1), and the difference between a political speech and a PI report (institutional factor in Experiment 2). It was necessary to verify that respondents in fact identified systematically between a positive and a negative evaluation (evaluation content factor in Experiment 2).

For this purpose, respondents in Experiment 2 were asked whether the evaluation (PI report or political speech) was a positive or negative evaluation of the Department of Health, using a ten-point scale. The measure of evaluation negativity was subjected to a 2 (institutional context)  $\times$  2 (evaluation content) between-subjects ANOVA. As expected, this analysis yielded a main effect of negativity,  $F = 10.563$ ,  $p = .001$ . Respondents who were presented with a negative evaluation judged the text to be more negative ( $M = 4.10$ ) than those who were presented with a positive evaluation ( $M = 4.98$ ). No other main effects or interactions were found. These findings confirm that respondents identified systematically between a positive and a negative evaluation.

### **Vignette reading check**

Further measures were used to address the risk of 'lazy' respondents who would answer the questions without reading the vignette, a risk that might be aggravated in the less controlled setting of web-based surveys. In the introduction to the questionnaire, respondents were told that the first set of questions following the 'event story' would be about certain facts in the story, for which they were asked to choose the 'correct' answer. These were simple 'factual questions' about the failure story (in Experiments 1 and 2), and the evaluations that followed (in Experiment 2), easily answered by a respondent who had read the vignette. The probability of getting all questions right by chance was .083 in Experiment 1, and .021 in Experiment 2. Respondents who erred in any of the questions were omitted from the sample, resulting in a total sample of 399 (a reduction of 16 per cent). This sample was used for the re-weighting process and analysis.

## **RESULTS**

### **Responsibility attribution**

A great deal of consistency was found across the responses to the various questions on responsibility attribution. Respondents who judged the minister/department/agency as responsible were more likely to see the need for resignation and reform. All eight items measuring responsibility attribution load highly on one dimension, and thus were used to produce a single index of overall responsibility attribution (based on principal component

factor analysis). These relationships attest to the construct validity of responsibility attribution measurement.

For the purpose of Experiment 1, separate indices were needed for the different hierarchical levels. This was done by averaging the items measuring each level separately. Based on experimental groups one and two together ( $N = 282$ ), subjecting these indices to within-subject ANOVA has shown that hierarchical level is a highly significant predictor of responsibility attribution ( $F = 106.633, p < .001$ ). Respondents judged the minister as less responsible ( $M = 6.05$ ) than the department ( $M = 7.11$ ), and the agency was judged as most responsible ( $M = 8.05$ ). Figure 2 depicts the relationships between hierarchical level and responsibility attributions for the three different measures ('responsibility', 'appropriateness of resignation' and 'need for reform').

**Experiment 1**

Hypothesis 1 predicts a reduction in the responsibility attributed to the appointing minister, compared with the responsibility attributed to the minister when no appointment is made. The data do not support this hypothesis. The two items assessing responsibility attribution to the minister were averaged ( $r = .68, p < .01$ ). The appointment had no significant effect on responsibility attribution in a two-tailed t-test ( $p = .95$ ). Separate testing of all 8 responsibility attribution items did not reveal any significant effect in any of the hierarchical levels.

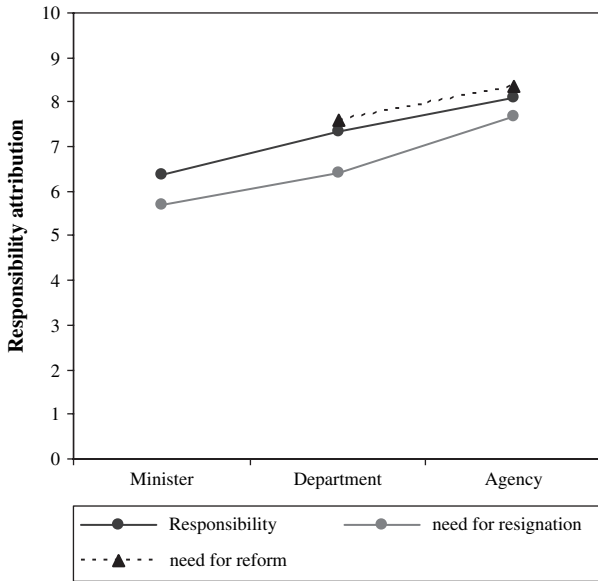


FIGURE 2 Responsibility attributions for different hierarchic levels

### Experiment 2

Hypothesis 2 predicts that an evaluation made by a public inquiry enjoys a higher level of credibility and, therefore, results in a greater effect on responsibility attribution than other political evaluations. The single index of overall responsibility attribution was subjected to a 2 (institutional context)  $\times$  2 (evaluation content) between-subjects ANOVA. This analysis yielded a main effect of evaluation content,  $F = 8.067$ ,  $p = .005$ , that is, respondents who received a positive evaluation attributed less responsibility ( $M = 6.97$ ) than those who received a negative evaluation ( $M = 7.69$ ). However, no significant main effect was found for the institutional factor and, most importantly for Hypothesis 2, no significant interaction effect was found for the two factors on responsibility attribution. Similar results were found in separate analyses of the three hierarchical dimensions. Hence, although the content of evaluation (both of PI reports and political speeches) affected responsibility attribution judgements, this effect was not significantly different between PI reports and political speeches – that is, public inquiry reports did not exert more influence on responsibility judgements than political speeches. Figure 3 depicts the effects of institutional factor and content factor on responsibility attribution.

### Credibility and accuracy

As mentioned above, Hypothesis 2 assumes a main effect of institutional factor on credibility; in other words, that a public inquiry report would be judged as more credible than political speeches. The item measuring credibility was subjected to a 2 (institutional context)  $\times$  2 (evaluation content)

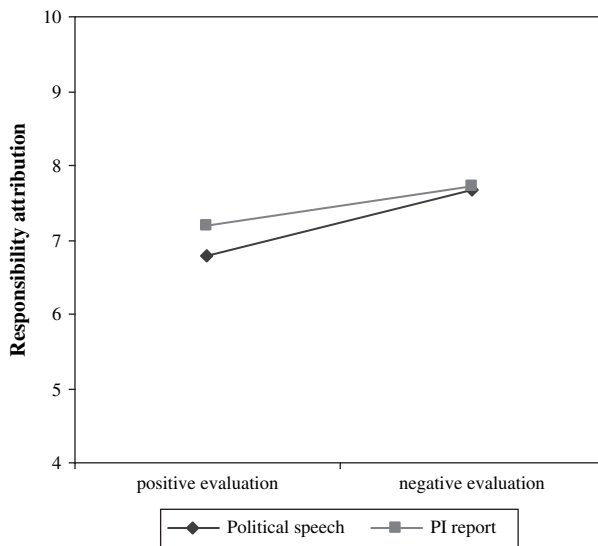


FIGURE 3 *Responsibility attribution effects*

between-subjects ANOVA. This analysis yielded a main effect of institution factor,  $F = 24.519$ ,  $p < .001$ , and a main effect of content factor,  $F = 13.267$ ,  $p < .001$ . However, an interaction effect was also found,  $F = 12.844$ ,  $p < .001$ . This highly significant interaction effect required reinterpretation of the main effects (Glenberg 1996, p. 402).

A closer look at the results reveals no difference in credibility judgements between positive and negative political evaluations, while a highly significant effect was found for PI reports. Respondents who received a positive political speech by the Minister of Health judged it to be roughly as credible ( $M = 3.85$ ) as a negative speech by the Head of the Opposition ( $M = 3.77$ ). However, respondents who received a positive PI report judged it as less credible ( $M = 4.13$ ) than respondents who received a negative PI report ( $M = 6.22$ ). Figure 4 depicts the interaction effect of institutional factor and content factor on credibility judgements.

The accuracy measure was subjected to a 2 (institutional context)  $\times$  2 (evaluation content) between-subjects ANOVA. This analysis yielded a main effect of institutional factor,  $F = 11.921$ ,  $p = .001$ , and a main effect of content factor,  $F = 34.285$ ,  $p < .001$ . No interaction effect was found. These findings suggest that negative evaluations were judged as more accurate ( $M = 6.92$ ) than positive evaluations ( $M = 5.43$ ), and PI evaluations were judged as more accurate ( $M = 6.68$ ) than political evaluations ( $M = 5.78$ ). These data confirm the expectation that the accuracy judgement would be more influenced by the evaluation content (a substantially greater effect) than by the institutional factor.

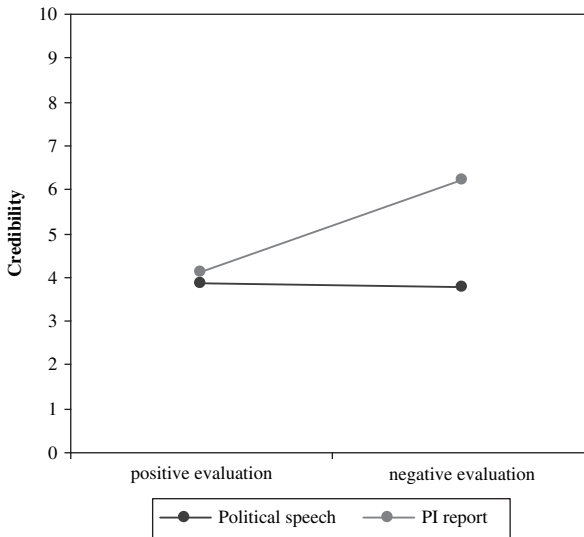


FIGURE 4 *Credibility effects*

### General trust in the judiciary

A number of writers have expressed concern that the involvement of senior judges in PIs into sensitive and contested political crises would undermine public trust in the judiciary (Drewry 1975, 1996; Lord Morris of Aberavon 2003). Thompson also addresses this question yet concludes that 'the judiciary's relationship with the public is unlikely to be adversely affected by these sorts of inquiries, as it will not trust anyone else to do them' (1997, p. 188).

The measure of general trust in the judiciary was based on a standard question drawn from the Mori Political Attitudes Poll, and the response was given on a ten-point scale ('For each type of person in the following list, please indicate how much you generally trust them to tell the truth or not?'). This item was subjected to a 2 (institutional context) x 2 (evaluation content) between-subjects ANOVA. This analysis yielded a main effect of institution factor,  $F = 5.182$ ,  $p = .024$ . However, these findings suggest that judges enjoy slightly higher levels of trust from respondents who had shortly before been presented with a PI report ( $M = 6.26$ ), than from those who read about the same affair without the involvement of a PI ( $M = 5.69$ ).

Considering the finding that the credibility of PI reports was dependent on evaluation content (above), this item was also subject to a two-tailed t-test, comparing the case of a positive PI report with that of a negative PI report. This analysis yielded no significant effect on general trust in the judiciary.

## DISCUSSION

The findings of these experiments do not lend support to the two hypotheses tested. The appointment of a PI does not seem to affect (or reduce) responsibility attribution to the appointing office holder (Hypothesis 1), and the general trust enjoyed by judges does not seem to increase the effect of PI reports on responsibility attribution, compared to textually equivalent public evaluations made by politicians (Hypothesis 2). The second experiment demonstrates that public evaluations did have an effect on responsibility attributions, yet the fact that the source of evaluation was a PI did not have a particularly greater influence on public perceptions of responsibility worthiness, compared with evaluations made by politicians.

The findings also demonstrate that respondents were systematically sensitive to institutional information in allocating responsibility. For all measures of responsibility attributions, respondents judged the agency as the most responsible, followed by the department, and ending with the minister, as depicted in figure 2, above. This distribution corresponds with the framing of the 'failure story' as located primarily in the agency. Although the experiment was not designed to test this particular aspect, it lends preliminary support to the feasibility of blame-avoidance 'agency strategies' (Hood 2002) in the British political system. These strategies are 'attempts by office holders or institutions to avoid or limit blame by the way that formal responsibility, competency or jurisdiction is allocated among institutions and office holders'

(Hood 2002, p. 16). This runs contrary to untested arguments that such strategies are not expected to be effective in shifting blame (Horn 1995, p. 46).

The findings of Experiment 2 suggest that PI reports enjoy relatively higher levels of credibility only when they are critical of the government. This greater acceptance of negative evaluations could be a manifestation of lasting trends in social characteristics, such as social mistrust in individuals in positions of authority or expertise or a more transient ebb in public trust. Determining this would require further research. Furthermore, in interpreting this finding, it should be noted that the particular 'failure story' used in the experiment prompted reactions which rendered a negative evaluation as a substantially more accurate representation of the affair. It is possible that a different story would have resulted in different acceptability of negative or positive evaluations. Thus, the findings should be understood to represent the credibility of public evaluations under conditions in which negative evaluations are judged by the public as more accurate than positive ones. Further replications of this experiment, using different experimental vignettes are, therefore, required.

While negative evaluations were judged as more accurate than positive evaluations, the difference in evaluation content did not affect the level of credibility ratings of political speeches, that is, there was no 'improvement' in credibility rating when a negative (more accurate) evaluation was given. This can be easily explained in the context of the political actors involved. Both are seen as having a stake in the evaluation they present (the minister in the positive evaluation and the opposition leader in the negative). On the other hand, one would expect that, given the higher level of trust in judges and their political independence, a judge chairing a PI would not be seen as having a stake in any particular outcome of the report; thus, we could have expected the same non-effect on credibility judgement, with both negative and positive PI reports getting relatively high credibility ratings (at about the level found for the negative PI evaluation). Yet the results were different. It appears that instead of a mechanism in which credibility reinforces the effects of evaluations content, the findings suggest that it is the content of evaluation which conditions credibility. I refer to this finding as 'conditional credibility' of public inquiries, since their credibility is dependent upon the content of their report.

A measure of caution is called for. As noted earlier, the experimental sample is not representative of the British population, and this raises the possibility that particular social characteristics may influence the findings to a degree that they would jeopardize the validity of generalization. A number of measures were employed in order to account for this risk and to enhance the external validity of the results. The experiment did not rely on a 'convenience sample' of undergraduates, as is often the case in such experiments. Preliminary tests of the data were conducted, and where correlations between social characteristics (gender and voting intentions) and the dependent variables were found, they were adjusted by re-weighting to have equal

proportions in experimental groups, and approximate these to the proportions in the general population. The results of the re-weighted sample were compared with an un-weighted sample. No significant differences were found. Finally, reports of earlier research suggests that social demographic characteristics do not have a significant impact on trust in political institutions (Levi and Stoker 2000); thus the risk in an unrepresentative sample appears to be low. As in any first experimental study of a social phenomenon, further research and replication is required to assess the robustness and external validity of the findings. In the present study, the case of the Hutton Inquiry was also analysed and compared to the experiment's results.

### **Conditional credibility in the Hutton Inquiry**

The findings of this experiment echo some of the developments in the 2003 Hutton Inquiry and its aftermath. Furthermore, the fact that the 'failure story' used in the experiment and the one underlying the Hutton Inquiry are so different provides additional credence to the attempt to generalize from this experiment.

As for the lack of a relative advantage in influencing public opinion by PI reports, Kellner (2004) has noted that, contrary to many predictions, voting intentions barely shifted at all in the wake of the Hutton Report, and there were only very slight increases in Tony Blair's satisfaction rating, as well as in the figures regarding the government as 'honest and trustworthy'.

Whatever Hutton said, too many people still thought Blair lied minus 45 per cent, which was only a three-point improvement on what they said before Hutton reported. (Toynbee 2004, pp. 363–4)

An even more striking similarity can be found regarding conditional credibility. Although strictly speaking, only an experiment design can allow the comparison of a scenario and its counter-factual, some comparative observations of perceptions before and after the Hutton Report are telling. The Hutton Report is widely perceived as an overall favourable evaluation of the government's handling of the David Kelly affair, and thus comparable to the 'positive' PI report scenario simulated in the experiment. Gamble and Wright note that:

[Hutton], hailed for his robust independence and forensic skills while his inquiry was sitting was transformed into an object of vilification and ridicule once his report appeared. (2004, p. 100)

Support for Gamble and Wright's observation of the widespread perception of the Hutton Inquiry as credible, as well as potentially detrimental to the Blair Government, prior to publication of the report can be found in abundance in the British press of that time.

Lord Hutton's stewardship of the investigation into Dr Kelly's death *has drawn praise from lawyers and politicians who regarded it as a template for future judicial inquiries.* (Verkaik 2003, italics added)

And:

Lord Hutton has alarmed the government by refusing to send drafts of his report into the death of David Kelly to ministers, officials and others, including the BBC, who will be the subject of criticism. His decision . . . could give Tony Blair only hours to react before the *potentially damaging report* is published. (Eaglesham 2003; italics added; see also Hughes 2003)

The 'perceptual U-turn' made almost immediately following the report's publication can be seen as a demonstration of the 'conditional credibility' finding. It is as if the two conditions tested in the experiment – exposure to a negative PI report, and exposure to a positive PI report – were emulated in the Hutton case by a sharp temporal division – before and after the report's publication. It is argued that until Lord Hutton's report was published, most (if not all) observers of the inquiry process assumed that the report was likely to be critical of the government's handling of the affair. Possibly the most prominent issues were the alleged political pressures over the editing of the 24 September 2002 Dossier (*Iraq's Weapons of Mass Destruction: The Assessment of the British Government*), and the Ministry of Defence's treatment of Dr Kelly. Under this impression, observers had thought they were in a situation simulated in the experiment by the experimental group which was exposed to a critical PI report. That group in the experiment rated the inquiry as credible and, as it appears, so did the observers and commentators of the Hutton Inquiry prior to the report. Once the report was publicized, the public and press had to accept that, in fact, the report was positive about the government's record – the situation simulated in the experiment by the group which was presented by a positive PI report. That group rated the inquiry report as much less credible – judging it to be as credible as a political speech. Similarly, in the Hutton case, the press and virtually all other commentators have adjusted their credibility rating accordingly.

Chairman: To complete the overview, when you were sitting you were sainted: you were this fearless forensic investigator?

Lord Hutton: Yes.

Chairman: The moment you reported, you were an establishment lackey?

Lord Hutton: Yes. (Public Administration Select Committee 2004)

This striking similarity, combined with the fact that the Kelly affair and the experimental 'failure story' represent such different situations, suggests that conditional credibility of public inquiries is potentially a more general finding than can be inferred by the experiment alone. As noted earlier, 'conditional credibility', as found in the experiment, is based on the different levels of credibility given to positive and negative PI reports. Gamble and Wright (2004) point to the great divide between the credibility enjoyed by the Hutton



Inquiry prior to its report, and the (general lack of) credibility it received after submitting it. If, indeed, the credibility change observed by Gamble and Wright can be understood in the context of 'conditional credibility', one could accept the notion that in the period prior to the report, public perceptions can maintain a high level of trust in the operation of the PI, and retain this trust if a negative report emerges. This perception is, however, abandoned if an 'overly' positive report is delivered.

### **A descriptive social role of public inquiries: from 'fact finding' to 'providing confirmation'**

These findings suggest not only a more sceptical view towards the independence and neutrality of judges while presiding as PI chairs (PASC 2005), but possibly help sketch a descriptive social role of PIs – one that is different from their normative role. The prevailing concept of PIs relies on the notion of fact finding. According to this notion, when a publicly distressing event occurs, an objective body is required to provide a factual account of the event. This notion implicitly relies on a pre-assumption that finding the facts would remedy an *a priori* state of factual ignorance. Under such a state the public is assumed to be ignorant of the crucial facts of the affair, and thus unable to establish sound responsibility judgements. More importantly, it is assumed that members of the public perceive their own state as such. In this situation it is reasonable to assume that a more credible investigator would be more influential on judgement formation than others (Hypothesis 2).

However, by relying on the findings of the experiment and the Hutton case, an alternative understanding of the social role of PIs is conceivable. In a society where mass media provides rich and immediate information (even if not always accurate), it is possible that people would rely on it to shape their opinions. Thus they would not see themselves as ignorant of the crucial facts. They may draw their judgements regarding responsibility well before an inquiry report is presented. Such a collective state of mind is supported by policy analysts' observations that in contemporary Western societies, there is no longer much mental and political space for accepting fatalist explanations like chance or act-of-god; adverse consequences are conceived as resulting from policy failures (Bovens and 't Hart 1996). Yet members of the public, almost as a reflex reaction, still seek a public inquiry in the aftermath of fiascoes and disasters. Immediately after the general disappointment from the Hutton Report (55 per cent considered it a whitewash; 26 per cent found it judicious and balanced), 54 per cent of the public thought there should be a judicial inquiry into the war in Iraq, while only 32 per cent thought there should not be one (YouGov Survey, 30–31 January 2004).

Under conditions where people generally want PIs after adverse events, while in the same time condition their credibility judgements of these *ad hoc* institutions on the inquiries' findings, we should consider an alternative to the 'fact finding' narrative on the social role of PIs, and consider a possibly more descriptive one – as 'providing confirmation'. According to this role,

people form value judgements in relatively early stages of an affair, and seek the appointment of a PI not (only) to reveal unknown facts, but to provide an authoritative confirmation of their *ex ante* judgements. Such a confirmation provides a source of empowerment. Seeking confirmation, not merely facts, can account for the apparent dynamics were the output (report) qualifies the process (inquiry), and not vice versa.

### **A modified explanation to the political decision to appoint a public inquiry**

In what way can these findings advance our understanding of the politics of PIs? If PIs do not offer responsibility attribution advantages in the short term, nor do they possess increased influence on public opinion, what thesis may provide an explanation to the political employment of these *ad hoc* institutions?

First, the employment of PIs in particular situations is the result of political decisions. Unless one subscribes to the assumption that politicians have perfect foresight of the consequences of political acts, one should not discard the possibility that politicians wrongly believe in the potential of the hypotheses tested. The prevalence of these hypotheses in the literature and their echo in 'common-sensical' claims by political players (see also Hattersley 2000, above) support this possibility. For this reason, the findings of this experiment do not refute the notion of 'borrowed authority' in so far as it is an explanation of the political decision to appoint a PI. It merely demonstrates the expected ineffectiveness of this notion as a mechanism for influencing public responsibility attribution.

A different contribution of these findings to the understanding of the politics of PIs may lie in suggesting an underlying mechanism for other (yet untested) arguments regarding blame-avoidance advantages of employing PIs. As noted above, the hypotheses tested in this experiment were limited to responsibility judgement strategies. Other arguments refer to agenda-control advantages in appointing PIs. However, these 'agenda-control' advantages do not offer a detailed explanation as to the reason *why* the appointment of a PI may provide a demotion of the inquired event on the public agenda. In other words, what does the government give in return for the time gained?

Based on the notion of 'providing confirmation', it is possible that the consent of political rivals, the press, and the public to demote the event on the public agenda is gained by the government's willingness to pay in the prospect of a negative authoritative evaluation. Political rivals, the press, victim groups and the general public may all be willing to agree to a pause, considering the prospect that their claims will be empowered by the critical report of a public inquiry appointed by the government. By the time this fails to materialize (and in the less frequent cases when it does), time has already been gained. This hypothetical mechanism incorporates 'conditional credibility' and elaborates the 'agenda-control' arguments in the literature. Further research is needed to test these modified arguments.

## ACKNOWLEDGEMENT

Funding for this research was provided by the Department of Politics and International Relations, Oxford University – Norman Chester Fund. Many people have read and commented on earlier drafts of this paper. I am particularly grateful to Christopher Hood, as well as to Orit Alfi, Daphna Canetti-Nisim, Sander Dekker, Robert Schwartz and Limor Sheffer. I'm also in debt to all the participants of the 'Blame avoidance and blame management' Workshop at the ECPR Joint Sessions 2005, especially Annika Brändström, Tereza Capelos, Will Jennings, Moshe Maor, Charles Raab and Kent Weaver. I owe a special thanks to Catherine Wreyford for her advice on preparing the vignettes. Finally, I thank the three anonymous referees for their helpful comments. Any faults that remain are of course my own.

## REFERENCES

- Arceneaux, K. 2003. 'The Conditional Impact of Blame Attribution on the Relationship between Economic Adversity and Turnout', *Political Research Quarterly*, 56, 1, 63–71.
- Arnold, R.D. 1990. *The Logic of Congressional Action*. New Haven, CT: Yale University Press.
- Baron, J. 2000. *Thinking and Deciding*, 3rd edn. Cambridge: Cambridge University Press.
- Baron, R.M. and D.A. Kenny. 1986. 'The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations', *Journal of Personality and Social Psychology*, 6, 1173–82.
- Berens, R.P., A.K. Bohara, J. Jenkins-Smith, et al. 2003. 'The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples', *Political Analysis*, 11, 1, 1–22.
- Bloom, H.S. and D.H. Price. 1975. 'Voter Response to Short-run Economic Conditions: The Asymmetric Effect of Prosperity and Recession', *American Political Science Review*, 69, 4, 1240–54.
- Bovens, M. and P. 't Hart. 1996. *Understanding Policy Fiascoes*. New Brunswick, NJ: Transaction Books.
- Bovens, M., P. 't Hart, S. Dekker and G. Verheuel. 1999. 'The Politics of Blame Avoidance: Defensive Tactics in a Dutch Crime-Fighting Fiasco', in H. Anheier (ed.), *When Things Go Wrong*. Thousand Oaks, CA: Sage, pp. 123–47.
- Brändström, A. and S. Kuipers. 2003. 'From "Normal Incidents" to Political Crises: Understanding the Selective Politicization of Policy Failures', *Government and Opposition*, 38, 3, 279–305.
- Clarke, Rt. Hon Lord Justice. 2000. *Thames Safety Inquiry: Final Report*. London: Cm 4558, HMSO.
- Department for Constitutional Affairs. 2004. *Effective Inquiries: A Consultation Paper produced by the Department of Constitutional Affairs*. London: CP 12/04, 6 May 2004.
- The Times*. 1988. Leading Article, 'Dissecting Tragedy', 13 December.
- Dowding, K. and W.T. Kang. 1998. 'Ministerial Resignations 1945–97', *Public Administration*, 76, 3, 411–29.
- Drewry, G. 1975. 'Judges and Political Inquiries: Harnessing a Myth?', *Political Studies*, 23, 1, 49–61.
- Drewry, G. 1996. 'Judicial Inquiries and Public Reassurance', *Public Law*, Autumn, 368–83.
- Eaglesham, J. 2003. 'Hutton Report "Will be bolt from blue"', *Financial Times*, 2 December, p. 1.
- Elliott, D. and M. McGuinness. 2002. 'Public Inquiry: Panacea or Placebo?', *Journal of Contingency and Crisis Management*, 10, 1, 14–25.
- Entman, R.M. 1989. 'How the Media Affect What People Think: An Information Processing Approach', *Journal of Politics*, 51, 2, 347–70.
- Entman, R.M. 1993. 'Framing: Toward Clarification of a Fractured Paradigm', *Journal of Communication*, 43, 4, 51–8.
- Fischhoff, B. 1975. 'Hindsight ≠ Foresight: The Effects of Outcome Knowledge on Judgement under Uncertainty', *Journal of Experimental Psychology*, 1, 288–99.
- Fischhoff, B. and R.L. Beyth. 1975. "'I knew it would happen" – Remembered Probabilities of Once-future Things', *Organizational Behavior and Human Performance*, 13, 1–13.
- Flinders, M. 2001. *The Politics of Accountability in the Modern State*. Burlington, VT: Ashgate.

- Gamble, A. and T. Wright. 2004. 'The Politics of Trust', *Political Quarterly*, 75, 2, 99–101.
- Gibson, J.L. and A. Gouws. 1999. 'Truth and Reconciliation in South Africa: Attributions of Blame and the Struggle over Apartheid', *American Political Science Review*, 93, 3, 501–17.
- Glenberg, A.M. 1996. *Learning from Data: An Introduction to Statistical Reasoning*. Mahwah, NJ: Lawrence Erlbaum.
- Griffith, J.A.G. 1997. *The Politics of the Judiciary*, 5th edn. London: Forntana.
- Hagiwara, S. 1992. 'The Concept of Responsibility and Determinants of Responsibility Judgment in the Japanese Context', *International Journal of Psychology*, 27, 2, 143–56.
- Haider-Markel, D.P. and M.R. Joslyn. 2001. 'Gun Policy, Opinion, Tragedy, and Blame-attribution: The Conditional Influence of Issue Frames', *Journal of Politics*, 63, 2, 520–43.
- Hart, H.L.A. 1968. *Punishment and Responsibility*. London: Oxford University Press.
- Hart, J. 1995. 'Presidential Lightning Rods – the Politics of Blame Avoidance – Ellis, R.J', *American Political Science Review*, 89, 4, 1021–2.
- Hattersley, R. 2000. 'The Diseased Herd', *Guardian*, October 28.
- Hood, C. 2002. 'The Risk Game and the Blame Game', *Government and Opposition*, 37, 1, 15–37.
- Horn, M.J. 1995. *The Political Economy of Public Administration: Institutional Choice in the Public Sector*. Cambridge: Cambridge University Press.
- House of Lords. 2003. Vol. 648 Part no. 97, 21 May 2003: Column 883.
- Howe, G. 1999. 'The Management of Public Inquiries', *Political Quarterly*, 70, 3, 294–304.
- Hughes, D. 2003. 'My job is on the line over top-up fees, admits Blair', *Daily Mail*, 3 December, p. 8.
- Iyengar, S. 1989. 'How Citizens Think About National Issues: A Matter of Responsibility', *American Journal of Political Science*, 33, 4, 878–900.
- Keeton, G.W. 1960. *Trial by Tribunal*. London: Museum Press.
- Kellner, P. 2004. 'Hutton, Ministers, the BBC and "Iraq Fatigue"', *YouGov* (<http://www.yougov.com/interactive/kellnerMain.asp?jID=3aId=994sID=6wID=0UID=>), accessed 4 February.
- Kingdon, J.W. 1995. *Agendas, Alternatives, and Public Policies*, 2nd edn. New York: Addison-Wesley/Longman.
- Kremnitzer, M. 1989. 'The Landau Commission Report – Was the Security Service Subordinated to the Law or the Law to the "Needs" of Security?', *Israel Law Review*, 23, 216–79.
- Lau, R.R. 1985. 'Two Explanations for Negativity Effects in Political Behavior', *American Journal of Political Science*, 29, 1, 119–38.
- Levi, M. and L. Stoker. 2000. 'Political Trust and Trustworthiness', *Annual Review of Political Science*, 3: 475–507.
- Lipsky, M. and D.J. Olson. 1977. *Commission Politics: The Processing of Racial Crisis in America*. New Brunswick, NJ: Transaction Books.
- McGraw, K.M. 1991. 'Managing Blame: An Experimental Test of the Effects of Political Accounts', *American Political Science Review*, 85, 4, 1133–57.
- McGraw, K.M. 1996. 'Political Methodology: Research Design and Experimental Methods', in R.E. Goodin and H.-D. Klingemann (eds), *A New Handbook of Political Science*. New York: Oxford University Press.
- McLean, M. 2001. 'How does an Inquiry Inquire? A Brief Note on the Working Methods of the Bristol Royal Infirmary Inquiry', *Journal of Law and Society*, 28, 4, 590–601.
- McLean, I. and M. Johns. 2000a. *Aberfan: Government & Disaster*. Cardiff: Welsh Academic Press.
- McLean, I. and M. Johns. 2000b. "'Regulation Run Mad": The Board of Trade and the Loss of the Titanic', *Public Administration*, 78, 4, 729–49.
- Polidano, C. 2001. 'An Exocet in a Red Box: Parliamentary Accountability in the Sandline Affair', *Public Administration*, 79, 2, 249–75.
- Public Administration Select Committee (PASC), House of Commons. 2004. *Government by Inquiry: Oral Evidence of Lord Hutton*, 13–14 May, HC-6061.
- Public Administration Select Committee (PASC), House of Commons. 2005. *Government by Inquiry: First Report of Session 2004–5, Vol. 1*, 3 February, HC-51-I.
- Rudolph, T.J. 2003. 'Who's Responsible for the Economy? The Formation and Consequences of Responsibility Attribution', *American Journal of Political Science*, 47, 4, 698–713.
- Salmon Commission of Tribunals of Inquiry. 1966. Cmnd. 3121.
- Segal, Z. 1984. 'Tribunals of Inquiry: A British Invention Ignored in Britain', *Public Law*, 206–14.

- Shadish, W.R., T.D. Cook and D.T. Campbell. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.
- Sulitzeanu-Kenan, R. and C. Hood. 2005. Blame Avoidance with Adjectives? Motivation, Opportunity, Activity and Outcome', paper presented at the *ECPR Joint Sessions*, Granada, April, 2005.
- Thompson, B. 1997. 'Conclusion: Judges as Trouble-Shooters', *Parliamentary Affairs*, 50, 1, 182–9.
- Timmins, N. 1985. 'Medicines watchdog sued over Opren: Victims of anti-arthritis drug to make claim on committee on safety of medicines', *The Times*, 23 July.
- Toynbee, P. 2004. 'The Trust Fund', in S. Roger (ed.), *The Hutton Inquiry and its Impact*. London: Politico's/Guardian Books.
- Verkaik, R. 2003. 'Hutton to quit as Law Lord after Kelly Report', *Independent*, 12 December, p. 5.
- Wade, H.W.R. and F.C. Forsyth. 1994. *Administrative Law*. 7th edn. Oxford: The Clarendon Press.
- Wheare, K.C. 1955. *Government by Committee*. Oxford: Oxford University Press.
- Weaver, R.K. 1986. 'The Politics of Blame Avoidance', *Journal of Public Policy*, 6, 4, 371–98.
- Woodhouse, D. 1995. 'Matrix Churchill: a Case Study in Judicial Inquiries', *Parliamentary Affairs*, 48, 24–39.
- Lord Woolf of Barnes. 2004. *Government by Inquiry: The Use of Investigatory Inquiries by Government*. Memorandum submitted to the Public Administration Select Committee. HC 51, GBI 22, Ev 183.

*Date received 17 January 2005. Date accepted 6 June 2005.*

**APPENDIX 1** *Sample characteristics (N = 474)*

Category		%
Gender	Female	64.0
	Male	36.0
Age	18–25	14.5
	26–30	16.2
	31–40	38.7
	41–50	17.7
	51–60	9.8
	61–65	2.3
	Over 65	0.8
Education	Less than high school	2.7
	Some high school	3.2
	Completed high school	34.8
	Some university	13.8
	Other tertiary education	17.8
	University degree	19.8
	MA degree	2.8
	Doctoral degree	2.2
Professional degree (MD, JD)	3.0	
Place of residence	East Midlands	7.8
	East	4.2
	London	8.7
	North east	6.3
	North west	13.7
	Northern Ireland	3.0
	Scotland	9.5
	South east	19.8
	Wales	4.7
	West Midlands	4.8
	Yorkshire	6.7
	Humber	0.3
Other	10.5	
Voting intentions	Labour	17.7
	Conservative	24.0
	Liberal Democrat	20.7
	Other	16.5
	Would not vote	21.0

**APPENDIX 2** *The original event***LAYOUT**

*The Times* (London), Tuesday 23 July 1985

'Medicines watchdog sued over Opren/Victims of anti-arthritis drug to make claim on committee on safety of medicines'

Nicholas Timmins, Social Services Correspondent

The Committee on the Safety of Medicines (CSM) is being sued by hundreds of people who claim to be victims of the anti-arthritis drug, Opren, which was withdrawn in 1982 after the deaths of 83 people and almost 4000 cases of side-effects.

In what is thought to be the first time the government's drug safety watchdog has been sued for damages, the Opren 'action committee' announced the action yesterday, which involves about 200 writs. The action committee claimed it had evidence of a 'comprehensive failure' by the Department of Health and the CSM 'to give the public the protection it needed and deserved to expect'. The writs, linking the CSM and health ministers as the licensing authority, with the manufacturers, Eli Lilly and Dista Products in claims for damages, have been issued, lawyers for the action committee said. More than 1000 claimants were expected to become involved.

Mr Jack Ashley, Labour MP for Stoke-on-Trent South, is demanding a public inquiry. In a letter to Mrs Margaret Thatcher, he said: 'The issues to be examined should include the relationships between members of the CSM and drug companies, the secrecy of the CSM deliberations, and criteria for granting product licenses and the efficacy of clinical trials'.

Opren was withdrawn within two years of its launch, but while the CSM granted the drug a licence, the action committee says it was refused in the United States, Sweden, Australia, Canada and New Zealand. Pre-launch trials involved 291 patients in the United Kingdom.

The group is also questioning the involvement of Professor Sir Abraham Goldberg, who joined the CSM shortly afterwards. According to the action committee, Sir Abraham was involved in trials on Opren before its launch. Both Sir Abraham and the Department of Health, which speaks for the CSM, declined to comment yesterday, saying the matter was *sub judice*.

If the action were successful, it is expected that damages would be met by the taxpayer through the Department of Health rather than by individual members of the committee, which is a statutory body appointed to advise ministers.

**APPENDIX 3** *The 'failure story' used in the experiment*

*This text was presented to the respondents of all experimental groups, and the only information presented to experimental group 1:*

In 2001, the new anti-arthritis drug Nerpo was approved by the Agency for the Safety of Medicines within the Department of Health. It was described, at the time, by the Minister of Health as 'an exciting breakthrough in the treatment of arthritis, which could greatly reduce the suffering of patients'. Three years later Nerpo was withdrawn from public use by the Department of Health after it was found that 83 people have died and almost 4000 suffered serious side-effects as a result of using the drug. However, Dr Clifford, a former pharmacologist at the Department of Health, disclosed on BBC News that the decision to withdraw the drug came almost a year after America's Surgeon-General had issued a public warning about using Nerpo. The Department of Health, together with the producer of Nerpo have been sued by more than 1000 claimants for damages. The Department may face damages payments of more than £ 80 ?£ million, at a time when the Health system is already facing financial difficulties.