
STUDIES IN REGULATION

On the discovery and assessment of
economic evidence in competition law

Chris Decker and George Yarrow

NS 1.1

2011

REGULATORY POLICY INSTITUTE

Chris Decker and George Yarrow

On the discovery and assessment of
economic evidence in competition law

Studies in Regulation, New Series, Vol. 1, No. 1

*Published by the Regulatory Policy Institute
2 Market Street, Oxford OX1 3ET, UK
www.rpieurope.org*

First published December 2011

© Christopher Decker and George Yarrow

1. Introduction

Assessment of economic data/information/evidence in more or less any context can raise difficult issues. Compared with the testing of hypotheses and theories in the physical sciences, there is much less ability to make use of controlled, repeated/repeatable experiments; and the resulting limitations are often reinforced by the complexity of the relationships and interactions that are involved in the determination of economic outcomes. This complexity necessarily leads to high degrees of uncertainty, particularly (though by no means exclusively) uncertainty about the future.

In the context of the application of public policy, the difficulties are multiplied by the inherent limitations of governmental processes for discovering and assessing economic information for decision-making purposes. Such limitations derive ultimately from two, interlinked, ‘monopolistic’ characteristics of government: a highly constrained ability to handle large quantities of potentially relevant information, and a lack of incentives to ‘get it right’.

It seems sensible, therefore, to entertain only relatively modest ambitions in relation to the potential for improved policy performance consequential on improved economic assessments. Yet even if the bar is set at a modest height, experience in the enforcement of competition law suggests a recurrent failure to rise to the challenge.

In what follows, we will seek first to understand why this is the case, and then to make some suggestions that might help mitigate the problems. No very comprehensive ‘solution’ can be expected, for reasons that are explored in section 2 below, which considers the nature of assessment problems. However, by focusing on a particular issue, ‘confirmation bias’ (in section 3), and on a particular area, competition law enforcement (in section 4), it is possible to identify some specific manifestations of the general problem with a degree of precision that opens up the possibility of remedial measures.

It can be noted at this point that the decision to focus on economic assessments made in the context of the enforcement of competition law should not be interpreted to imply that the problems under discussion are abnormally severe in this particular area of government activity. In fact, our view is almost the opposite: competition law enforcement appears to us to be where some of the highest quality (micro-)economic assessment work has been done within government, an outcome that we would attribute in part to the appeals mechanisms that have been established in this area, which serve to ensure that economic reasoning is subject to ‘competitive’ challenge. If then, there are significant weaknesses in economic assessment procedures in competition law enforcement, we invite the inference that it is

likely that substantially greater problems are to be found in other parts of government activity.

In any given context not all types of evidence and information will have equal value for the purposes at hand. We therefore explain and discuss the crucially important notion of the *diagnosticity* of evidence in section 5, where we seek to outline a framework for thinking about strategies for information discovery and assessment. The utility of this framework is explored in section 6, which illustrates its capabilities by using it to identify and explain two assessment failures and one assessment success in three past competition policy cases. Section 7 contains some brief, concluding comments.

2. The nature of the problems

Key to the understanding of what economic analysis can (and can't) contribute to public policy development is recognition that what usually stands to be assessed is the operation of *complex, adaptive systems* of social interactions.¹ A market is an example of such a system, although, in practice, it is better viewed as a particular sub-system within a wider system of interacting markets.

The 'systems' focus is particularly clear at the beginnings of political economy, even where it takes rather different forms; as in Quesnay's *Tableau Économique*² or Smith's *Wealth of Nations*. It is reflected, for example, in the emphasis on unintended consequences (the *Invisible Hand* metaphor being an example) and on the indirect or unseen effects of public policy, as opposed to the direct or more visible impacts.³ The systems focus is retained in modern economic analysis in the form of 'models' – which are built up from sets of propositions, hypotheses and assumptions about agents and their interactions – that are simplified representations of the system of interest (although this later, modern analysis tends, all too often, toward over-abstraction/over-simplification in general, and toward abstraction from the adaptive characteristics of actual economic systems in particular, usually in the name of mathematical tractability).⁴

¹ A system refers to a set of interacting agents, in which the structure and nature of the connections/relationships between agents is important in determining behaviour and outcomes. Adaptation refers to the capability of a system to change (e.g. in its structure, and hence in its behaviour and performance) in response to external perturbations or shocks, and to internally generated changes, particularly the discovery of new information. Complexity is used to denote a situation in which the information determining how the system behaves and evolves is beyond the comprehension of any single agent (whether inside or outside the system).

² François Quesnay (1759), *Tableau Économique*.

³ See also Frédéric Bastiat (1850), *What is seen and what is not seen*.

⁴ The dominant aesthetic in economics is to prefer precise and definite answers to questions posed, even where they are wrong, to answers that are more nuanced and roughly right.

Notwithstanding the contributions of econometrics to the developments of techniques aimed at addressing some of these limitations, the wide variations in economic contexts across markets, across locations, and over time, tend to imply that detailed, case-study approaches – such as are to be found in the best work of competition authorities – generally offer the most promising approach to disentangling and understanding those *systems* of interactions that attract the label ‘economic’. At a more general level of understanding – for example, of what Adam Smith called the ‘connecting principles’ of political economy – this process is aided by the building up of case studies conducted in a wide variety of different contexts. As Smith’s friend and colleague David Hume put it:

“Mankind are so much the same, in all times and places, that history informs us of nothing new or strange in this particular. Its chief use is only to discover the constant and universal principles of human nature, by showing men in all varieties of circumstances and situations, and furnishing us with materials from which we may form our observations and become acquainted with the regular springs of human action and behaviour. These records of wars, intrigues, factions, and revolutions, are so many collections of experiments, by which the politician or moral philosopher fixes the principles of his science, in the same manner as the physician or natural philosopher becomes acquainted with the nature of plants, minerals, and other external objects, by the experiments which he forms concerning them.”⁵

Case study work is much more easily described than accomplished, however. What typically faces the analyst is a complex, factual reality, coupled with relatively disordered sets and sources of information relating to that reality. Given this, it can be asked: what principles should guide the collection and assessment of economic evidence? For example, in relation to information gathering there are various questions about the types of information to be sought, and about when to stop the information search process (when has sufficient information has been collected for the purposes at hand?). In relation to subsequent assessment of the available information, there are potential questions about: the type and number of hypotheses to be developed and explored; whether the available information is such that classical statistical techniques can reasonably be used; whether the specific and/or limited nature of the evidence calls for Bayesian methods of evaluation; and so on.⁶

2.1 Sources of systematic errors (biases)

It should be clear from these and similar questions that there can be substantial scope for discretionary choices in the approach to evidence in particular cases, and further that these discretionary choices open up the possibility of approaches that may be influenced by a range of factors other than a disinterested search for truth, or for best possible understandings of the matters at hand within the relevant resource limits (search typically requires time and

⁵ D Hume (1748), *An Enquiry Concerning Human Understanding*,

⁶ Bayesian methods provide an inductive framework for revising prior beliefs in response to new, relevant information.

money). Such influences include individual psychological factors, organisational and social pressures, and material interests in outcomes that might be affected by the results of assessments. A brief outline of some of these sources of potential bias illuminates the depth of the problems to be faced.

Individual psychology

The study of individual decision making has identified what has become a rather long list of possible cognitive biases, although we note that the significance of these for the effectiveness of decision making is a matter of dispute; for example, bias *per se* may not be a problem if it is relatively small (choices are approximately right), if it is associated with quick, low-cost assessment processes, or if unbiased alternatives might be expected to lead to larger errors on average (i.e. they exhibit more variability). Perhaps the most salient contribution is the distinction between simple, fast heuristics and slower, more reflective approaches to assessment – what Kahneman has called ‘System 1’ and ‘System 2’ processes. This distinction has analogues at the social/institutional level which we believe might be helpful in thinking about policy making processes (see further below).⁷

In the psychological and behavioural economics research, fast (System 1) heuristics are found to exhibit systematic assessment errors in particular circumstances. This is hardly surprising: the relevant approaches have evolved or developed to allow for appropriate responses in particular situations/circumstances, and it is not to be expected that they will necessarily work well in other situations/circumstances. When new and difficult problems arise, the human mind tends to resort to slower, more reflective ways of responding to the problems posed (System 2), but this too has its limitations. More specifically, this type of assessment requires *attention*, which appears, on the evidence, to be rather limited in supply, and hence, in economic terms, *costly*. Speaking roughly, attention appears to both restricted in its field of vision and difficult/costly to sustain.

The implications of this research for assessments of complex economic systems are obvious. The behaviour of systems is a function of a large number of factors. Fast heuristics tend to focus on easily visible, simple, direct connections between specific, often localised, actions and easily visible, often localised, consequences of those actions. That is, in Bastiat’s terminology, they focus on ‘what can be seen’ (quickly); and systematic error can be expected as a result.

Slower, more reflective thinking is required to address the less direct, more diffuse, less visible (often unintended) consequences of human conduct. However, in a complex situation, attention devoted to particular aspects of the situation can easily lead to neglect of

⁷ D Kahneman (2011) *Thinking, Fast and Slow* (Penguin, London).

information concerning other aspects of the situation that may be of critical importance for understanding the behaviour of the system as a whole.⁸

In practice, the problems here are compounded by the tendency of much economic theorising to rely heavily on systematic abstraction (i.e. to neglect large numbers of potentially relevant factors in the name of building solvable models). This gives rise to what might be called the ‘Quine problem’, which might be exemplified by considering the assessment of a hypothetical ‘theory of harm’ in a competition law case.

Quine⁹, following Duhem¹⁰, argued that tests of theories are, in effect, joint tests of the propositions that, taken together, make up the theories: they are not tests of individual propositions or of sub-sets of the propositions. If a particular theory is found to be at odds with reality, in the sense that its implications are found to be inconsistent with available evidence, it is not possible to assign that failure to the failure of a particular proposition within it. Thus, a component proposition might be true, yet the theory may fail as an explanation of what has happened; or a component proposition might be false, yet the theory as a whole can still work.

This is, in effect, an argument against reductionism – the approach of breaking things down into components and assessing those components one by one – which fits easily with a similar argument that appears in the analysis of complex, adaptive systems, to the effect that, to understand behaviour, systems must be analysed as a whole. Unfortunately, even with slow thinking, there are challenging problems in ensuring that, under the required, holistic approach, *all* the factors that may be material for the purposes at hand have been given an adequate allocation of (scarce) *attention* – and *attention* here refers to something much more than simply listing a particular feature or aspect of the situation that has been noticed and recorded.

If, then, what is actually assessed is a joint hypothesis or theory, made up of several (potentially many) elements or components, there is necessarily scope for holding on to favoured, component propositions, even when the theory has been found to be inconsistent with available evidence. This can be accomplished by changing other components of the relevant theory in ways that, combined with the retained, favoured proposition, move the

⁸ In *Thinking, Fast and Slow*, Kahneman gives the example of an experiment by Christopher Chabris and Daniel Simons, who made a short film of two teams passing basketballs, one team in white the other in black. Viewers were asked to count the number of passes made by members wearing white shirts, and to ignore the other players. Halfway through the video, a woman wearing a gorilla suit appears for about nine seconds, crossing the court, thumping her chest, and then moving on. About half the people who view the video report, on completion, that they have seen nothing unusual. The implication is that in focusing attention on one aspect of what is happening the viewer may be blind to other things.

⁹ W.V.O. Quine (1951) “Two dogmas of empiricism”, *The Philosophical Review*. Vol 60, pages 20-43.

¹⁰ P Duhem (1954) *The Aim and Structure of Physical Theory*, trans. Philip P. Wiener (Princeton, NJ: Princeton. University Press).

implications of the theory toward a closer relationship with the evidence. Just as data can be mined to produce a desired outcome, so the component elements of theories can be mined to ‘protect’ a favoured implication.

Thus, although a holistic approach can be said to rest on sophisticated, reflective (System 2) thinking, it too can be associated with considerable inertia in conclusions in the face of challenging evidence. In economics, if there is a ‘demand’ for a bias in interpretation of evidence, or for putting a favoured implication, inference or belief less at risk from the evidence, there are typically many ways in which that demand can be met, as anyone familiar with the use of cost-benefit analysis in such circumstances will likely have found.¹¹

Organisational influences

Where the relevant assessments are made within an organisational context – which is the case for those issues in competition law that are decided by the administrative agencies of UK/EU competition law/policy – approaches taken to economic evidence can also be expected to reflect features of the relevant bodies. Public organisations have their own agendas and priorities; and they develop cultures in response to challenges of adapting to their own, changing social/political/environments and of internally integrating the individuals and sub-groups who make up the organisation.

Most regulatory assessments are undertaken in organisations that are strongly influenced by political and administrative/bureaucratic cultures, which may not be conducive to approaches to economic assessment aimed at understanding how markets actually work and how market behaviour might change when subject to perturbations. Administrative/bureaucratic cultures may give undue attention to internal processes (e.g. production of documents is seen as the priority) relative to external matters (e.g. how markets are changing) – other than those ‘semi-external’ processes (external to the organisation but internal to government) that determine the responsibilities, powers and budgets of different departments and agencies. In contrast, political cultures tend nowadays to be heavily pre-occupied on that which can be seen, and hence easily explained to the media and public (e.g. on a System 1 basis), to the neglect of the diffuse, unseen effects of decisions.

Private interests

In addition to the influence of organisational interests in survival and expansion and of the cultures that develop to promote such outcomes, public decisions can also be affected by the private interests of those participating in the decision making process and of those likely to be affected by decisions. A well known example from the extensive historical literature on

¹¹ Quine adopts the metaphor of a ‘web of belief’, where strands of the web are the individual elements or components, and where the strands connecting with the structure to which the web is attached are the most basic statements of fact.

regulation is the influence that is referred to as ‘regulatory capture’. Thus, regulators seeking ‘rewards’ in the form, say, of remunerative employment following the relinquishing of their public duties, or seeking political honours or political recognition, may be inappropriately influenced by agendas other than that of impartially evaluating evidence and reaching appropriate conclusions.

Such effects can occur at varying levels of an organisation, and are by no means confined to the senior executive level. Thus, the life of a hard-pressed middle-ranking official can be made easier by the co-operation of companies providing information and evidence. The moral sentiment of *reciprocity* can then lead to undue weight/attention being given to the submitted evidence which, in all likelihood, will be partial to at least some degree; and, in simply describing this kind of situation, it is almost self-evident that it is extremely difficult to eliminate the tendency completely.

2.2 *Starting from the beginning*

The surge of interest in the study of *heuristics and biases* over recent decades, now imported into economic thinking in the form of what is known as behavioural economics, is, in at least one sense, a return to the roots of the subject. Smith’s first major work was *The Theory of Moral Sentiments* (1759) – a book that can be characterised as a work of ‘philosophical psychology’ – which opens with, and is founded on, the notion of *sympathy*, hypothesised to be a universal aspect of human nature.

Critical to the reasoning in the *TMS* is the notion of the *impartial spectator*, who acts as a touchstone for the refinement of moral sentiments and the moderation of emotions, providing a guide as to what are and what are not appropriate sentiments in the various circumstances in which we may find ourselves. This notion has also been the subject of a recent revival, appearing as an important concept in Amartya Sen’s *The Idea of Justice*.

The basic idea is that accurate adjudication requires stepping away from one’s own material and emotional interests, and adopting a position that allows for the evaluation of as much *contextual* information as possible; recognising, at the same time, that this notion of objectivity should not be taken so far as to suggest an endorsement of some Archimedean point that is free of all biases.

Interestingly, and a point that is crucial for the later development of our arguments, in Smith’s thinking the recommended attitude with which the impartial spectator should approach judgments about human conduct is not one of general scepticism, which tends to be incompatible with *sympathy*. In the dialogues between actors and spectators (who will make the adjudications)¹², the spectator is expected to do what he/she can to understand possible

¹² The actor and the spectator can be the same person.

arguments about the conduct in question. At the same time there is a reciprocal responsibility on the actor, which is to provide relevant information as clearly as possible, and not to obfuscate or mislead. These mutual obligations imply a rudimentary *community of inquiry*¹³, and an associated *culture of inquiry*. Using this terminology, it can be said that, in this paper, one of our central concerns is with factors that tend to undermine the development and evolution of such communities/cultures of inquiry, and with ways of thinking about economic evidence that might help to foster such communities/cultures.

Since, consistent with Hume, the understanding of conduct requires careful attention to variations in context, accurate adjudication will generally only be possible on the basis of detailed information about the relevant circumstances. This puts a premium on information discovery, and it is a factor that favours sympathy over scepticism in approaches to context, at least in the earlier stages of enquiry (scepticism is less effective in teasing out the information on which later, effective adjudication will depend). It also points to another tendency that can lead to biases associated with failures in information gathering and analysis: excessive haste in preliminary adjudication.

Such excessive haste goes under many names. Psychologists refer to *premature closure* – a failure to consider, or to continue considering, reasonable alternatives after an initial view has been reached – which is, for example, recognised to be a major cognitive problem in medical diagnosis. More everyday labels include ‘jumping to conclusions’ and prejudice. Whatever the name, however, the point is that a sympathetic *impartial assessor* may be better at avoiding the bias than a more sceptical assessor.

At the end stages of an enquiry, of course, rejection of at least some propositions is to be expected. Indeed evidence kills theories (which economics spawns in great abundance), and stronger evidence will have more killing power than weaker. It is at these later stages that a more sceptical approach on the part of the assessor will tend to become warranted. First, however, the evidence has to be discovered and examined, and premature closure gets in the way of that.

The role of the impartial spectator in the *TMS* is chiefly to make evaluations of conduct in everyday situations. Moreover, since Smith placed emphasis on the motivations for conduct, rather than the effects of conduct, no great analytical capacity is typically required to reach conclusions. When evaluating the behaviour of complex systems, however, something more than this is required. Specifically, the evaluator needs to be *skilled*, as well as impartial; and also to be *engaged*, in the sense of being willing to bear the costs of the attention that will be required to perform a demanding task. Further, because attention may tend to focus on a

¹³ This is a term first used by C.S. Peirce to refer to the way in which people create knowledge in collaboration with others, and later taken up and generalised in the theory of education. Here the usage is much closer to the narrower, original conception of a group of collaborating investigators/scientists/scholars sharing similar values concerning the construction of knowledge.

limited number of aspects of a complex system and thereby lead to the neglect of other, relevant factors, effective economic assessment may often require the combined attention of several *spectators/evaluators*.

In administrative processes, the *skilled, engaged, impartial assessor* can be interpreted as an ‘idealised’ civil servant or regulatory official, or a group of such officials; in judicial processes, the relevant assessors are judges or tribunals. These (administrative and judicial assessment procedures) are different ways of doing things, and the different institutional structures can be expected to have different effects. In examining these matters, we suggest that the notions of an effective *community/culture of inquiry* can be of considerable assistance.

A central aspect of our general argument is that failures to create such communities/cultures – failures that manifest themselves in an excessive propensity to neglect, dismiss, discount, or not use reasonable endeavours to discover, relevant information – is a central weakness of many, current, institutions of competition law and regulation; but that the extent of the failure varies among types of institution. Perfection is impossible, but some forms of arrangements are less bad than others; and, by studying and understanding the source of the resistances to consideration of relevant evidence, we will be in a better position to develop and sustain the desirable *communities and cultures of inquiry*.

With these points in mind, we will examine, in the later parts of this paper, some of the issues that have arisen in EU and UK competition law over recent years in connection with the way in which competition authorities have gone about assessing economic evidence. In doing so, it is perhaps worth re-stating our view that this is unlikely to be the area of public policy where the effects of biases in the assessment of evidence are likely to be greatest. It does, however, have the advantage of being a policy area where some of the underlying issues with which we are concerned have been recognised and discussed, including by thoughtful senior public officials participating in the relevant assessment processes.

3. A core problem: confirmation bias

As already indicated, there has been increasing interest in recent years in the implications of psychological research on decision making for the conduct of public policy. The ‘headline’ stories concern the interest of politicians and policy makers in using the insights of this work to modify or affect the behaviour of the public in various areas of social and economic life. In contrast, our interest is in the potential use of the research to modify or affect the behaviour of decision-makers so as to mitigate some of the most egregious, systematic errors/biases in decision-making. More specifically, in relation to the enforcement of competition law, our focus is on biases in judgment in circumstances where the investigative, assessment and decision-making functions are all performed by the same organisation. This

focus leads fairly directly to consideration of the notion of *confirmation bias*, a concept that has a reasonably long pedigree in psychological literature, and which has been used extensively in contexts where different pieces of information/data are used to build or test a case or hypothesis (such as in medical diagnosis).

3.1 What is confirmation bias?

The term ‘confirmation bias’ has been used in slightly different ways in different parts of the relevant literature, but it generally refers to an approach to information gathering and decision making in conditions of uncertainty which is based on the inappropriate bolstering of particular hypotheses or beliefs. As Nickerson observes:

*“People tend to seek information that they consider supportive of favored hypotheses or existing beliefs and to interpret information in ways that are partial to those hypotheses or beliefs. Conversely, they tend not to seek and perhaps even to avoid information that would be considered counterindicative with respect to those hypotheses or beliefs and supportive of alternative possibilities.”*¹⁴

Put slightly differently, the bias can be said to arise through “*unwitting selectivity in the acquisition and use of evidence*” in assessment. Such selectivity ‘protects’ favoured or privileged hypotheses and beliefs from the threat of being undermined by new information/evidence. Given that we are here focusing on confirmation bias in a competition policy context, it might also be said that the tendency to selectivity in the approach to evidence ‘distorts competition’ among competing, alternative hypotheses; and this alternative language can be helpful when it comes to designing mitigation strategies.

There are a number of distinctions to be borne in mind when considering use of the notion of confirmation bias. First, at least at the level of the individual decision-maker, the concept, as developed in psychology, is generally not intended to refer to the purposive, deliberate or conscious decision to be selective in the use of evidence. Rather, the notion is defined broadly to capture a less conscious, less deliberate and in some cases unintentional approach to evidence and information gathering and assessment, which is nevertheless ‘one-sided’ in its outlook.

Second, this boundary between purposeful and unwitting behaviour becomes much fuzzier when dealing with organisational assessment and decision making, and confirmation bias then tends to become part of a wider pathology of organisational behaviours. Thus, from our own experience, there is often an ‘appreciation’ of the existence of bias and selectivity on the part of at least some parts of decision making teams, which is nevertheless suppressed for reasons that are sometimes conscious and sometimes not. This may be why some of those

¹⁴ R.S. Nickerson (1998) ‘Confirmation Bias: A Ubiquitous Phenomenon in Many Guises’, *Review of General Psychology*, vol. 2, number 2, 175-220.

who have been most vociferous in expressing concerns about the existence and effects of the bias in the competition policy area have often been public officials with direct experience of the relevant decision making processes.

Third, and different from the purposeful/unwitting distinction, is the distinction between interested and disinterested approaches to evidence, and this takes us back to the ideal of the skilled, engaged, impartial assessor. As discussed above, individuals and organisations can have material stakes in the outcomes of assessments, or can have particular sets of beliefs or ways of doing things (i.e. aspects of organisational cultures) that can become sources of discomfort when disturbed or challenged. Under the influence of such ‘partial’ interests, there can therefore be tendencies, on the parts of both individuals and groups, to believe what is convenient or useful to believe and to see what is convenient or useful to see – a sort of ‘partial’ or ‘private’ pragmatism (“this idea is acceptable on account of its manifest utility” (to the private or partial interest)), which may be compared the ‘public’ pragmatism of the scholars of that particular philosophical school.¹⁵

Of significance in the case for the use of the concept of the *impartial spectator* is the implicit assumption that partiality is a major, distorting factor in evaluation of evidence; although that doesn’t mean that it is the only significant factor. There is a range of decision-making contexts where an individual has no obvious, material, personal interest in a particular outcome or in maintaining a particular hypothesis – medical diagnosis and judicial decision-making might be given as examples – but in which systematic biases in assessment have been shown, in some circumstances, to exist. According to Nickerson, there is a ‘great deal’ of empirical evidence that supports the claim that confirmation bias is ubiquitous in a range of different decision making contexts, and that it appears in many guises. He cites areas such as government policy-making,¹⁶ medical diagnosis, and science as examples of where confirmation bias has been shown to exist, and to influence behaviour in terms of information gathering, analysis and decision-making.

Of course, there is a possible argument that, in fact, there is no such thing as the disinterested assessor, and that, for example, the evidence of confirmation bias in medical diagnostics is a reflection of the influence of the private interests of the doctors concerned. Thus, even in

¹⁵ C.S. Peirce, who introduced the notion of a *community of inquiry*, was one of the founders of Pragmatism as a modern philosophical school. The development of the notion may therefore have been a response to a perceived danger that a pragmatic approach to things could, under the influence of private/partial interests, tend to lead to biased discovery processes and distorted conclusions. If so, our arguments in this paper are following a part beaten track.

¹⁶ Nickerson cites Barbara Tuchman: *Wooden headedness, the source of self deception, is a factor that plays a remarkably large role in government. It consists in assessing a situation in terms of preconceived fixed notions while ignoring or rejecting any contrary signs. It is acting according to wish while not allowing oneself to be deflected by the facts. It is epitomized in a historian's statement about Philip II of Spain, the surpassing wooden head of all sovereigns: "no experience of the failure of his policy could shake his belief in its essential excellence."* Tuchman, B. W. (1984). *The march of folly: From Troy to Vietnam*. New York: Ballantine Books.

health systems where doctors have no direct financial interest in the outcomes of diagnosis, the *process* of diagnosis involves resource allocation decisions to be made by the assessor, whose own resources are limited. At what point, for example, should the doctor stop asking questions and searching for further information before settling on a view about the likelihoods of alternative causes of symptoms? It is not difficult to imagine circumstances in which such a decision might be influenced by ‘private (to the assessors) agendas’.

Similar points apply to assessors dealing with competition and regulatory matters. There may be no financial interest at stake, but the exploration of alternative competing hypotheses requires time and effort. Hicks said that ‘the best form of monopoly profit is a quiet life’, and a similar remark might be applied to the ‘perks’ (rather than profits) of administrators and adjudicators operating within systems that are, after all, typically highly monopolistic.

It can be easy, therefore, for busy assessors to be over-dismissive of inconvenient hypotheses, arguments and information; and this can result in a bias in administrative systems towards the reaching of premature judgments. This is *premature closure*; and subsequent confirmation bias might be seen as a way of sustaining such premature judgments: *"Once a policy has been adopted and implemented, all subsequent activity becomes an effort to justify it."*¹⁷

Given that judgments have to be made in conditions of limited resources, including limited time, early judgments can, of course, be ‘justified’ in terms of ‘administrative convenience or necessity’. Indeed, for many decisions, the justification will be convincing: this is not dissimilar to the System 1 and System 2 distinction in the psychological research, in that many simple decisions can effectively be taken without much effort. Here, however, we are concerned with assessments of complex systems.

It is perhaps ironic that one of the most common ways in which confirmation bias is sustained is by the dismissal of evidence on the basis that the relevant information has been provided by an interested party; a case of the pot calling the kettle black. This tends to be a feature of administrative/investigative/inquisitorial processes, but not of adversarial processes; which rely heavily on ‘competition’ as a discovery procedure: *"Truth reveals itself in the crucible of vigorous exchanges among those with competing perspectives."*¹⁸

As noted in passing earlier, and at least for practical purposes, impartiality is not an all-or-nothing matter. In the words of an Old Bailey¹⁹ Judge: *"We at the old Bailey play it straight down the middle, not too partial – not too impartial either"*.²⁰ If the influence of partial and private interests on assessments can be reduced, then we might expect that bias will be

¹⁷ Tuchman, *ibid.*

¹⁸ W.W. Park (2010), “Arbitrators and Accuracy”, *Journal of International Dispute Settlement*, Vol 1, No 1.

¹⁹ The Central Criminal Court of England and Wales.

²⁰ Cited by Christopher Bellamy (2006) “The Competition Appeal Tribunal – five years on”, in Colin Robinson (ed.), *Regulating Utilities and Promoting Competition: Lessons for the Future*, Edward Elgar Publishing.

reduced. However, it cannot be concluded that confirmation bias would necessarily tend to zero: it could still be the case that bias remains, consequential on other factors, such as the way in which the human brain processes information.

3.2 *Private and partial interests*

Developing social systems that are populated with individuals and organisations motivated by ‘partial and private interests’ but which operate in ways that are effective in serving the general/public interest has been a central concern of economics since its first emergence as a field of study. Smith summarised the issue in his *invisible hand* metaphor, when referring to a situation in which the unintended consequences of the pursuit of partial/private interests are favourable to general/public interests. Like Hume and other contemporaries, however, he was conscious of the fragility of such systems, and of the constant threat of unravelling posed by the unchecked pursuit of private and partial interests of varying forms.

As already indicated, the concept of partial and private interests is to be construed widely, and not confined to simple and direct financial or material interests in the outcomes of decisions. In relation to our current focus of attention, for example, it seems clear from both experience and more systematic research that the individual career motivations of practitioners – both lawyers and economists – can be important determinants of which competition law cases are pursued, and how they are prosecuted. Weaver, for example, in her study of the US Department of Justice Antitrust Division, found that a significant number of attorneys joined the Department because they wanted to ‘*bring cases and win them*’, and that, when a case does emerge, the “*staff will be energetic in seeking to develop it into an actual case, and it will be quite unabashedly partisan in its use of arguments*”.²¹ Katzmann observed a similar trend in his study of the Federal Trade Commission, concluding that choices as to which cases to pursue was, in part, the result of a struggle to pursue actions in the public interest while at the same time reflecting the ‘*decision making calculus*’ of commission staff including the need to satisfy staff morale and the professional objectives of staff attorneys (in terms of future job prospects).²² Finally, and more recently, Leaver examines the behaviour of US Public Utility Commission officials and finds that reputational and career concerns – particularly a desire to avoid criticism or excite interest groups – can bias bureaucratic decision-making. Leaver argues that an implication of this finding is that the appointment of academics, for example, to senior posts within an agency on a short-term basis may not bring the desired benefit of independence and neutrality, as the academics may “*also be driven by a strong desire to maintain peer respect, not to mention future jobs.*”²³

²¹ S Weaver (1977) *Decision to prosecute: Organization and Public Policy in the Antitrust Division* (The MIT Press Cambridge Mass.) pages 8 and 9.

²² RA Katzmann (1980) *Regulatory Bureaucracy; The Federal Trade Commission and Antitrust Policy* (The MIT Press Cambridge Mass.), pages 84 and 184-185.

²³ C Leaver (2007) ‘Bureaucratic minimal squawk behavior: Theory and Evidence from regulatory agencies’ Oxford University Department of Economics Working Paper number 344, August 2007.

3.3 The ‘scientific method’ and falsificationist approaches

The question of how individuals collect and use information to form judgements in the context of uncertainty and imperfect information is one that has long occupied the minds of philosophers²⁴ and psychologists.²⁵ It has been an issue of particular interest to those who are interested in the history and philosophy of science, including the social sciences.

An important proposition that has been explored, and empirically tested, in some of this work is the extent to which individuals tend to adopt a so-called ‘scientific approach’ to information gathering and analysis, particularly in contexts where they are acting dispassionately and independently. This approach, which is often associated with the work of the philosopher Karl Popper, is based on the argument that the only way to establish the truthfulness, or accuracy, of a proposition or hypothesis is to seek out information which could falsify the proposition or hypothesis.

Popper’s *falsificationist* position has been challenged by other philosophers, but of greater relevance here are the numerous experimental and empirical studies that have shown that this is not generally how decision makers – including professional investigators such as scientists and doctors – tend to approach evidence gathering and hypothesis testing. Thus, in line with what has been said above, a much more frequently adopted approach appears to be biased toward seeking out information that tends to confirm a ‘privileged’ rule or hypothesis and to not examine information that may challenge or contradict a prevailing or favoured hypothesis. As Oswald and Grosjean observe “*To be blunt... humans do not try at all to test their hypotheses critically but rather to confirm them.*”²⁶

This, however, may be too bleak a view of things. One possible response is to treat Popper’s approach as normative, as a standard to be aspired to, rather than as a description of practice, even among professional adjudicators; although we note that this does not deal with what many, since Hume, have regarded as the unavoidable, practical case for inductive reasoning (drawing necessarily uncertain conclusions from relatively limited information), which exists notwithstanding the logical arguments for scepticism. Another is to recognise that, when there are competing hypotheses, the confirmation of one is logically connected to the rejection of others. That is, the issue is less one of setting out to confirm or reject a particular

²⁴ Francis Bacon, *Novum organum*, reprinted in Burt, E. A. (Ed.), *The English philosophers from Bacon to Mill* (pp. 24-123). New York: Random House.

²⁵ Thurstone, L. L (1924). *The nature of intelligence*. London: Routledge & Kegan Paul (page 101).

²⁶ ME Oswald, and S Grosjean S.(2004), Confirmation bias, in *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Hove, UK: Psychology Press (citing Wason, P. C. (1960). ‘On the failure to eliminate hypothesis in a conceptual task’ *Quarterly Journal of Experimental Psychology*, 14, 129-140).

hypothesis, but of recognising the set of linked relationships between evidence and alternative, competing hypotheses and of approaching assessment in a way that reflects these links. We will argue below that Bayesian inference, which can be regarded as a form of inductive logic, provides an appropriate framework for empirical assessment.

3.4 A positive test strategy approach

At this point in the discussion, and again to suggest that the Oswald and Grosjean view may be too pessimistic, it is useful to introduce a further distinction between the notions of confirmation bias and what is called a positive test strategy (PTS). Klayman and Ha have defined a PTS as an approach to hypothesis testing which only examines instances in which:

- (1) a property or event is expected to be observed (to test if it does occur) or
- (2) by examining instances in which it is known to have occurred (to see if the hypothesised conditions prevail).²⁷

In short, a PTS seeks to provide a heuristic that narrows and restricts the information space to instances or examples that have some prior probability of being relevant.²⁸

The point here is that, in circumstances where it can be expected that there will be a significant number of instances/observations that are potentially capable of falsifying hypotheses, a more efficient approach to information gathering is to restrict the field of inquiry (whilst taking care to ensure that the restriction does not unduly limit the number of falsification opportunities). Oswald and Grosjean describe the advantage of this approach as follows: “A PTS would enable us to find the ‘needle’ much more easily since we would not have to search the whole ‘haystack’”.²⁹

A PTS, therefore, is something different from an attempt to confirm hypotheses or immunise them against rejection. A PTS consciously approaches information gathering and hypothesis testing by examining *only* information or applying only tests that are capable of confirming the hypothesis (but simultaneously capable of rejecting it). It is necessarily, and deliberately, restrictive in its focus. This differs from an approach where no prior decision is made to restrict the information collected or tested to assess the validity of a hypothesis, but where the information that is collected during the inquiry tends to have the characteristic that it will

²⁷ J Klayman and YW Ha (1987) ‘Confirmation, Disconfirmation and Information in Hypothesis Testing’ *Psychological Review*, vol 94, page 212.

²⁸ ME Oswald, and S Grosjean, S.(2004), Confirmation bias, in *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 79–96). Hove, UK: Psychology Press pages 81- 82.

²⁹ *Ibid*, page 82

likely confirm the desired hypothesis, and/or will be unlikely to be inconsistent with that hypothesis (ie: confirmation bias).

While there is debate as to the precise nature of the distinction between a PTS and confirmation bias, in our view, the critical difference appears to be that a PTS approach involves a *conscious* and deliberate decision to narrow the information examined when testing a hypothesis, whilst taking care to preserve the possibility of observations that would imply rejection of the hypothesis; whereas approaches manifesting confirmation bias do not consciously adopt such a strategy *a priori*. Rather, as noted above, confirmation bias is more of a subconscious, non-deliberate cognitive strategy, based on seeking out only information consistent with a preferred hypothesis, or on discounting or neglecting information that may challenge the preferred hypothesis.

In section 5 below, we develop further our own general position on these points, which is one that is distinct from advocating either a strict falsificationist approach or, alternatively, a positive testing strategy. Rather our suggested approach embraces the notion of *diagnosticity*³⁰, and therefore adopts an approach to information gathering that seeks to differentiate between a number of plausible, competing hypotheses. This leads to a focus on tests that provide the maximum information gain or impact with respect to beliefs (see the discussion of diagnosticity below), rather than simply trying to falsify or confirm a particular hypothesis. The key idea is that a given piece of evidence can have simultaneous implications for a number of different hypotheses, and that this (obvious) point should be borne constantly in mind in the processes of discovering and evaluating information.

4. Confirmation bias and competition law enforcement

The discussion in the preceding section has centred on the issue of confirmation bias in a general sense, noting that the issue can arise in a range of different contexts and settings where there may be a cognitive or organisational tendency to build a case to justify a particular conclusion or to sustain a specific hypothesis. Returning to the central focus of this study, however, it is also clear that the issue of confirmation bias is one that is potentially of major significance in the specific context of the enforcement of competition law.

4.1 Economic evidence and competition law enforcement

³⁰ A notion that has similarities with Popper's notion of the *severity* of a test of a hypothesis. See F.H. Poletiek (2009), "Popper's *Severity of Test* as an intuitive probabilistic model of hypothesis testing", *Behavioral and Brain Sciences*, 32: pp 99-100.

Recent decades have seen a significant shift in the role and prominence given to the assessment of economic evidence, information and data in competition law investigations in the European Union. An important aspect of this general shift has been a movement away from what has been described as a formalistic, *per se* type approach to competition enforcement based on a so-called ‘more economic’ approach, centred around ‘effects-based’ analysis.³¹

While generally heralded as a positive development by economists and other commentators, this shift in enforcement practice has brought with it a set of practical complications. In particular, it has created the challenge of developing an appropriate framework for identifying, collecting and assessing relevant economic evidence in competition law proceedings. Among other things, this has involved asking questions such as: What are the relevant substantive economic questions/issues that should be asked/addressed in a particular case? What economic evidence is relevant to an assessment of those questions? How, practically, should assessors go about analysing and sifting through the different pieces of economic evidence (including data) submitted? How should assessors respond to different pieces of evidence that appear to be of equal probative weight/value but that present conflicting signs/indications of whether there is a competition problem?

As already implied, these issues are not unique to competition law enforcement. Similar sorts of questions arise in many other areas of diagnostic activity where a decision-maker is required to identify, collect and impartially analyse different pieces of information, and then to assess the implications of such information for various alternative or competing hypotheses or propositions (medicine is an obvious example, but areas such as intelligence gathering are also comparable).

Recent judgments of the European Court have served to highlight some of the problems associated with the use of economic evidence in practice. One problem identified is that the enforcement agency (in this case the European Commission) is being confronted with ever-increasing amounts of economic and/or econometric evidence during the administrative procedure. This raises the very practical issue of how to sift through the evidence presented and decide the relative weights to be given to different pieces of evidence (such as potentially conflicting economic/econometric analysis). This specific issue was raised in *GlaxoSmithKline Services Unlimited v Commission (GSK)*, an action under Article 81 (now 101) of the EC Treaty, where six separate expert reports were submitted to the Commission during the administrative procedure.

³¹ For example see the roundtable discussion in ‘Economics Experts before Authorities and Courts Roundtable’ in B Hawk (ed) *International Antitrust Law & Policy: Fordham Corporate Law 2004* (Juris Publishing New York 2006) 621-622; N Kroes ‘Preliminary Thoughts on Policy Review of Article 82’ (Speech at the Fordham Corporate Law Institute New York, 23rd September 2005); Commission of the European Communities Antitrust: consumer welfare at heart of Commission fight against abuses by dominant undertakings Press Release IP/08/1877 (3 December 2008).

On appeal the Court of First Instance (CFI) was critical of the Commission's handling of this evidence, concluding that the decision was "*vitiated by a failure to carry out a proper examination, as the Commission did not validly take into account all the factual arguments and the evidence pertinently submitted by GSK, did not refute certain of those arguments even though they were sufficiently relevant and substantiated to require a response, and did not substantiate to the requisite legal standard its conclusion that it was not proved.*"³²

Whereas *GSK* highlights a recurring criticism of administrative authorities to the effect that they have failed to take relevant factors into account (i.e. there have been errors of omission), another area of difficulty has been the sufficiency of the reasoning presented by the Commission as to why it has chosen to rely on one piece of economic evidence or information in preference to another. This has led to the claim that the Commission has been somewhat selective (either consciously or unconsciously) in its treatment of economic evidence. For example, in reviewing the Commission's decision to prohibit the merger between *Airtours/Firstchoice*, the CFI was critical of the Commission's interpretation of specific pieces of economic data in relation to demand growth, observing that this was '*inaccurate in its disregard*' for evidence which showed considerable market growth, and that the Commission '*failed to produce any more specific evidence*', and as such '*was not entitled to conclude that market development was characterised by low growth*'.

A similar issue arose in *Commission v Tetra Laval BV (Tetra Laval)* where the CFI concluded that the "*evidence produced by the Commission was unfounded*" and "*was unconvincing*" and that the Commission's forecast "*was inconsistent with the undisputed figures ... contained in the other reports*".³³ On appeal, the European Court of Justice (ECJ) affirmed the CFI's judgement noting that the Commission's conclusions "*seemed to it to be inaccurate in that they were based on insufficient, incomplete, insignificant and inconsistent evidence*".³⁴

A related question that has faced the Commission concerns the scope and scale of the evidence that it might reasonably be expected or required to collect and assess in any investigation; which is a matter of some significance in the context of significant constraints on the resources that are available for enforcement. In recent decisions, the European Court appears to have established a relatively high standard in respect of: (1) the scope of evidence gathering that should be undertaken, and (2) how the Commission should then analyse and assess that evidence once collected. The ECJ has noted, for example, that in complex and difficult merger cases, the Commission's approach requires "*a particularly close examination of the circumstances which are relevant for an assessment of that effect on the*

³² Case T-168/01 *GlaxoSmithKline Services Unlimited v Commission of the European Communities*. [2006] ECR II-02969, para [303].

³³ Case C-12/03 *Commission of the European Communities v Tetra Laval BV* [2005] ECR I-00987, para [46].

³⁴ *Ibid*, para [48].

conditions of competition in the reference market". Furthermore, the ECJ noted that in order to adduce proof of anti-competitive effects in such complex cases the Commission needed to undertake "*a precise examination, supported by convincing evidence, of the circumstances which allegedly produce those effects*".³⁵

In other decisions, the European Court has stressed that the Commission needs to assess whether the factual arguments, and evidence upon which they are based, are "*relevant, reliable and credible*",³⁶ and that the relevant standard is that the Commission adduce "*precise and coherent evidence demonstrating convincingly the existence of the facts constituting those infringements*".³⁷ More specifically, the Court has noted that in the context of both mergers and investigations under Article 102 the evidence needs to be "*factually accurate, reliable and consistent*" and that the "*evidence contains all the information which must be taken into account in order to assess a complex situation and whether it is capable of substantiating the conclusions drawn from it.*"³⁸

We note at this point that, although some have argued that the Court is now setting an unduly high standard, this criticism is, in our view, unconvincing. The ECJ's judgments reflect nothing more than what Smith and Hume would have regarded as the *minimum necessary* attention to detailed, contextual information if an economic approach is to have any value at all in the enforcement of competition law. Particularly given the propensity of many economists to engage in over-abstract analysis, strict requirements that attention be paid to factual detail and to all significant factors is essential if sound conclusions are to be drawn in circumstances where the effects of business conduct – which is what the economic assessments are mostly concerned with – tend to be so heavily influenced by contextual factors.

We also note that the problems are by no means confined to the European Commission. UK enforcement agencies, for example, have been similarly chided in recent years by the Competition Appeals Tribunal; but the underlying problems – and, in particular, the tendency of the administrative agencies to over-economise on the use of contextual evidence – have been around for a long time. Thus, in perhaps the most famous minority report of its history, following an inquiry into *The Supply of Beer*, a Member of the Monopolies and Mergers Commission, Mr Leif Mills, said of the conclusions and proposals of his colleagues on that case that:

³⁵ Ibid, para [24].

³⁶ Case T-168/01 *GlaxoSmithKline Services Unlimited v Commission of the European Communities*. [2006] ECR II-02969, para [263]

³⁷ Ibid, para [7].

³⁸ Case C-12/03 *Commission v Tetra Laval* [2005] ECR I-987, para [39]; Case T-201/04 *Microsoft v Commission* [2007] ECR II-3601, para [89].

*“Of course, there is a theoretical attraction in making major changes to the tied estate and the loan tie. However, the attraction smacks a little of the academic question ‘the brewing industry may well work in practice, but does it work in theory?’. The proposals seem designed to fit the structure of the industry into some sort of theoretical Procrustean bed.”*³⁹

As might be hoped, and possibly even expected, these and other criticisms have stimulated responses from the authorities. The assessment procedures of the UK Competition Commission (the successor to the MMC) have been radically overhauled since the time of the Beer investigation (1986-9), and there have been a number of changes at the European Commission aimed at improving the Commission’s handling of economic evidence and information. Among the more significant changes was the establishment of the Office of the Chief Economist, and the appointment of a Chief Economist. In addition the Commission has published a ‘best practice’ guideline for dealing with economic evidence in competition law matters.⁴⁰ An important component of this guidance is the need for a coherent and ‘testable hypothesis’ (or more broadly a ‘case theory’) as part of any enforcement strategy. This at least encourages the explicit checking of the theory against available data and information, although it falls somewhat short of addressing the kinds of issues identified by Quine.

4.2 *The potential for confirmation bias in competition law enforcement*

Enforcement of EC competition law is based on an administrative/ inquisitorial model where the same body/organisation is responsible for functions such as the decision to ‘open’ a case, the investigation of that case, and reaching a final decision on the evidence (which then forms the basis for any recommendation to the College of Commissioners). In these circumstances, where various functions are bundled or combined within a single organisation – and in some cases undertaken by the same team of individuals – there would appear to be few checks and balances on the development of cognitive and organisational biases (either intentional or unintentional); biases that are capable of affecting the evidence that is collected and its evaluation.

That, given the administrative model, confirmation bias is likely to be a particular problem in EC competition law enforcement has been recognised by current and former senior staff members of the European Commission. Wils, for example, identified confirmation bias as one of the possible causes of a prosecutorial bias in EC antitrust enforcement noting that:⁴¹

³⁹ Monopolies and Mergers Commission (1989) *The Supply of Beer – A report on the supply of beer for retail sale in the United Kingdom*, para [43].

⁴⁰ European Commission (2010) *Best practices for the submission of economic evidence and data collection in cases concerning the application of Articles 101 and 102 TFEU and in merger cases* Directorate General of Competition. See also the guidelines of the UK Competition Commission (2009) *Suggested best practice for submissions of technical economic analysis from parties to the Competition Commission*.

⁴¹ W Wils (2004) ‘The Combination of the Investigative and Prosecutorial Function and the Adjudicative Function in EC Antitrust Enforcement: A Legal and Economic Analysis’ *World Competition*, vol 27, p 216.

“If such confirmation bias is indeed a general tendency of human reasoning, there is no obvious reason why the persons within the European Commission dealing with an antitrust case would be immune from it. The question is however why they would start from an initial belief that there is an antitrust violation. In merger cases, there is no obvious reason why they would do so. ...The situation may be different with regard to Articles 81 and 82 EC where, certainly after the abolition of the notification system by Regulation No 1/2003,³⁹ an investigation will normally be started only if the officials from DG Competition hold the initial belief that an antitrust violation is likely to be found.”

Neven also discusses the potential for a prosecutorial bias in EC competition law enforcement, and notes a tendency for selectivity in the Commission’s processes:

“The observations that the Commission may decide early on cases and search for selective evidence or that theories are neglected are consistent with the incentives generated by the inquisitorial regime with a prosecutorial bias implemented by the EU.”⁴²

Moreover, he concludes that this tendency “to focus on one side of the evidence” might be exacerbated by various institutional factors in the European context:

“The prosecutorial bias and the intrinsic features of an inquisitorial procedure would appear to reinforce each other; in particular the conclusion that an inquisitor might not invest in seeking evidence towards both sides of the argument (when evidence is hard to manipulate) or might suppress conflicting evidence will be reinforced in the presence of a hindsight bias. The tendency towards extremism in the EU is also probably reinforced by the inconsistency between the standard of proof and the scope of the decisions mentioned above, at least with respect to the implementation of the merger regulation. Indeed, when evidence is not very conclusive, the Commission cannot meet the required standard of proof with either decision. In those circumstances, it will have a further incentive to shift towards extreme outcomes (by suppressing evidence or failing to fully consider some alternatives).”⁴³

The issues are not restricted to Brussels, however. Laura Carstensen, a Deputy Chairman of the UK Competition Commission, raised similar concerns. Speaking of the proposed merger between the Competition Commission and the Office of Fair Trading in the UK, she has said.

“And there are of course obvious potential disadvantages to merger, predominantly concerned with the possible impact on fairness from confirmation bias when the whole decision-making pathway in a case—from receipt of complaint to opening an initial investigation to choice of ‘tool’ to phase 1 decision to phase 2 decision and finally to

Wils was at the time a member of European Commission Legal Service. The article was, however, written in a private capacity.

⁴² D J Neven (2006) ‘Competition economics and antitrust in Europe’ *Economic Policy*, p 32. Professor Neven has been the Chief Economist at DG Competition, and we note that his reference to ‘early’ decisions/judgments in cases accords with our own experience, in regulatory evaluations as well as in competition law cases.

⁴³ *Ibid*, page 30.

remedies decision—takes place within a single body.”⁴⁴

More generally still, the UK National Audit Office, in its work on the evaluation of the quality of regulatory impact assessments,⁴⁵ has been strongly critical of government departments’ use of these assessments to ‘justify’ decisions that have already been taken or judgments that have already been made; rather than to use them as exercises in information discovery and analysis, the results of which are intended to *inform* the decision makers concerned. The motivation to ‘justify’ therefore turns the (costly) assessments into exercises in confirmation bias, particularly when there is an injunction to make the notional assessments ‘evidence based’. In practice, this simply means that the authors of the documents search for confirmatory evidence, adding costs without in any way affecting decision-making outcomes.

4.3 Examples of possible confirmation bias from recent competition cases

The general concerns about confirmation bias in EC competition law enforcement expressed above do not appear to be simply theoretical or academic speculation. For example, there appear to be a number of cases where the investigative and decision making processes of the Commission has been shown to be vulnerable to patterns which – to an outside observer – closely resemble or approximate confirmation bias, and which therefore indicate something beyond the occurrence of random administrative errors and failures (such as might be said to have occurred in at least some of the cases noted in section 4.1 above). Recent cases that have been appealed to the European Court provide support for this position.

The European Commission’s *Airtours* decision is perhaps the most obvious, first example. In that case, the Commission was found on appeal to have:

- selectively omitted certain elements among the facts available to it (such as evidence provided on the Form CO relating to the rate of demand growth);⁴⁶
- claimed reliance on a market study which, upon investigation by the Court, turned out to be reliance on a one-page extract;⁴⁷ and
- failed to produce an econometric study upon which it alleged it had relied in its decision.⁴⁸

⁴⁴ L Carstensen (2011) ‘Keynote speech by Laura Carstensen, Deputy Chairman of the Competition Commission, to the Association of Corporate Counsel Europe Seminar’, 9 March 2011.

⁴⁵ We note that there is a strong similarity between regulatory impact assessment and the evaluation of ‘remedies’ in some competition policy cases.

⁴⁶ M de la Mano (2008) ‘A theory on the use of economic evidence in competition policy cases’ Presentation at the Association of Competition Economists, Budapest, 28 November 2008

⁴⁷ Case T-342/99 *Airtours plc v Commission of the European Communities* [2002] ECR II-02585 para [128].

In another decision, *Impala v Commission*, the CFI was highly critical of the Commission's approach to the economic evidence, describing the economic assessment in that case as 'succinct, 'superficial', 'unsubstantiated' and 'purely formal', noting at one point that the Commission's treatment of campaign discounts (an important element of the case) was "imprecise, unsupported, and indeed contradicted by other observations in the Decision". In addition, the CFI was highly critical of the Commission's interpretation of the responses it received to questionnaires sent out to retailers, concluding that the Commission had misrepresented the number of responses received, and further that the statements did not support the conclusion the Commission drew from them.⁴⁹

More recently, and in the context of an Article 82 (102) investigation, Intel complained to the European Ombudsman regarding a number of alleged procedural and administrative errors committed by the Commission during the process of its investigations. In this matter, it was alleged that the Commission had failed to consider and take adequate account of potentially exculpatory information; specifically notes of a meeting between Commission officials and an important customer of Intel (Dell). In his judgement of July 2009, the European Ombudsman first noted the following standard for information gathering and how this interacts with the margin of appreciation/discretion afforded the Commission in complex cases:

*"While the Commission has a reasonable margin of discretion as regards its evaluation of what constitutes a relevant fact, the Commission, when seeking to ascertain relevant facts, should not make a distinction between evidence which may indicate that an undertaking has infringed Article 81 EC or Article 82 EC (inculpatory evidence) and evidence which may indicate that an undertaking has not infringed Article 81 EC or Article 82 EC (exculpatory evidence). In sum, the Commission has a duty to remain independent, objective and impartial when gathering relevant information in the context of the exercise of its investigatory powers pursuant to Article 81 EC and 82 EC."*⁵⁰

On the facts in this case, the Ombudsman concluded that the Commission had committed an instance of maladministration in failing to record details and a proper record of a meeting, which could have concerned evidence that was potentially exculpatory of Intel.

In sum then, it would appear that issues relating to confirmation bias are very real in the European context, and we will give further, illustrative examples in section 6 below. This view is shared by Neven, who has observed that: "Overall, it would thus appear that the self confirming biases that may be induced by the prosecutorial role that the Commission assumes cannot be dismissed as insignificant".

⁴⁸ Ibid para [132].

⁴⁹ Case T-464/04 *Independent Music Publishers and Labels Association (Impala, International Association) v Commission of the European Communities* [2006] ECR II-02289, paras [385]-[387].

⁵⁰ P Nikiforos Diamandouros (2009) 'Decision of the European Ombudsman closing his inquiry into complaint 1935/2008/FOR against the European Commission'. 14 July 2009, para [82].

And we repeat again that the problems are not confined to the EU level (we could equally well cite cases from UK domestic competition law), to administrative/inquisitorial arrangements (the difficulties are more general than that), or to competition law (we have seen much more severe problems in regulatory impact assessments).

4.4 Procedural suggestions for dealing with confirmation bias

Various proposals and suggestions have been presented as to how to address concerns regarding potential confirmation bias in the context of EC competition law enforcement. Wils, for example, suggests that the risks of confirmation bias could be reduced if there were better internal checks and balances, such as a second team of officials within the Commission to review the work of the first team of case handlers (e.g. a peer review team). Neven lists a range of other possibilities including a shift to a more adversarial system, which he argues would open up economic arguments to more detailed scrutiny, and would address incentives in inquisitorial systems to suppress economic evidence⁵¹ and to fail to consider alternative interpretations of economic evidence.⁵² Alongside this proposal he suggests increasing resources for the Commission to allow it to properly validate economic evidence, and changing the standard of review within the inquisitorial procedure, to improve decision-making. In the UK, Carstensen has suggested that institutionally separating the choice of cases, including initial investigatory work necessarily connected with that task, from the function of deciding cases can eliminate one of the major sources of confirmation bias.⁵³

These are all suggestions worthy of more detailed evaluation. The common themes are familiar ones, concerning the importance of checks and balances, and of competition or contestability in processes for discovering what is or has been the case. They also resonate both with wider liberal political values and with the economic notion of competition as a discovery process. It is encouraging that these ideas are being promulgated by people with first-hand experience of the problems of assessment, and, on the basis of our own experience, there can be little doubt that recognition of the existence of these problems is far more advanced in competition law enforcement than it is in other areas (such as regulatory impact assessments), with the possible exception of issues surrounding the evaluation of remedies in a sub-set of competition policy cases, most notably market investigation cases in the UK, where remedies have sometimes been highly intrusive and, in effect, have become a form of regulation (a reality that tends frequently to be denied, not least because, since much economic regulation is anti-competitive in its effects, acknowledgement of it implies

⁵¹ D J Neven (2006) “Competition economics and antitrust in Europe”, *Economic Policy*, pp 5-6, 30.

⁵² Neven notes for example that: ‘Evidence which is not subject to rigorous scrutiny can be easily abused: key assumptions in theoretical reasoning can be disguised as innocuous and empirical results that are not robust can be disguised as such’ Ibid, page 31.

⁵³ L Carstensen (2011) ‘Keynote speech by Laura Carstensen, Deputy Chairman of the Competition Commission, to the Association of Corporate Counsel Europe Seminar’, 9 March 2011.

recognition of the fact that competition authorities are, in practice, themselves capable of acting in anti-competitive ways).

Notwithstanding these positive developments, we think that there are more questions to be resolved about the wellsprings of biases in assessment processes before some of the more detailed aspects of such prospective reforms can themselves be assessed (i.e. there is a risk of ‘premature closure’/confirmation bias in the process of evaluating alternative assessment and decision processes). For example, one very specific ‘device’ that is sometimes used in group decision processes is that of the devil’s advocate. *Prima facie*, such a procedure appears meritorious as a method of countering obvious biases, including confirmation bias and groupthink, but in practice the results from its use have often been disappointing. The question then is: why?

Possible explanatory factors have been discussed, in a preliminary way, earlier on in this study: the relative merits of *scepticism* and *sympathy* on the part of skilled, impartial assessors (scepticism, such as may be embodied in a devil’s advocate, may be unwarranted until the end stages); and the merits of developing *communities of inquiry* (Satan’s advocate does not sound like a member of the community). At an organisational level, to the extent to which a case team is responsible for ‘building the case’ and maintaining a file, the ability of any ‘second pair of eyes’ to review what is and what isn’t in the file may be limited in complex cases (the information set is likely to be large); and the very fact of placing the different tasks within the same organisation arguably introduces a different, but powerful bias: *conformity bias*. When the ‘second pair of eyes’ (including those of a devil’s advocate) belong to the same organisation as those whose conduct or performance they are assessing, attainment of impartiality can be expected to be that much more difficult. Thus, devil’s advocacy may become a ‘tick-box’ exercise, with any real devilry missing from the process.⁵⁴

In sum, we suggest that it is appropriate to return to some rather fundamental (and unsettled) questions about the discovery and assessment of information for public policy decisions, before looking at potential ways forward. In the useful terminology of regulatory impact assessment, there is still work to be done on ‘identifying the problem’, before alternative ‘options’ are fully developed, and, *a fortiori*, before the expected effects of implementing those options can be assessed. In doing so, another centrally important concept will necessarily come to the fore, the *value (or usefulness) of information*. An alternative way of expressing the general problems examined in this paper is to say that there are systematic tendencies in public policy assessments to *under-value* certain types of available or potentially available information/evidence, and to *over-value* other types of information, at

⁵⁴ A more devilish suggestion is the pre-mortem (see G Klein, “Performing a project premortem”, *Harvard Business Review*, September 2007). This involves asking a group of people, who are knowledgeable about a particular piece of prospective policymaking, to imagine that the policy measures have been introduced and that they have turned out to be a disaster, and, on this basis, to write a short account of the history of the disaster. For example, what went wrong, when, why, and with what consequences?

least when valuations are made in relation to the ‘public’ objectives that the relevant decision makers are intended to pursue.⁵⁵

5. A diagnostic framework for evaluating economic evidence

Information may acquire value in a number of different ways. For example, information that is protected by a patent may have commercial value by virtue of the fact that the patent holder enjoys certain monopoly rights over its use.

In the current context, our interest lies in information that could be useful to a *skilled, engaged, impartial assessor* in making judgments about the behaviour of complex economic systems. Given this, how then might we go about thinking about the ‘value’ or usefulness of a particular piece of evidence or set of information?

5.1 *The single/privileged hypothesis approach*

An obvious starting point is to think about the contribution of evidence to evaluating the likelihood of a hypothesis of interest, and at this point it will be useful to distinguish between two types of ‘hypotheses’ that may be relevant. The first occurs where the hypothesis is a specific proposition or assumption, such as might be a building block of a wider theory. The second is where the hypothesis is an implication of a theory or model, such as a theory of harm in a competition law case.⁵⁶ We will refer to the first as an assumed hypothesis, and the second as an implied hypothesis.

Clearly, testing the second type of hypothesis is an exercise in evaluating the implications of a set of propositions/assumptions taken jointly (in the theory or model); but, in practice, even assessing simple propositions – which might be considered to be an exercise in fact-finding – may require the adoption of auxiliary propositions/assumptions, whether explicitly or tacitly. Consider, for example, the proposition that *the market price elasticity of demand is independent of price* (i.e. the market demand curve is iso-elastic or, equivalently, the relationship between quantity demanded and price is log-linear). To assess this proposition it may be necessary to undertake some applied econometrics, and this type of exercise will generally involve the adoption of a whole raft of further technical assumptions, without which it would be impossible to make inferences about demand elasticities from available data.

⁵⁵ Information will be valued differently when assessed on the basis of ‘private’ or ‘partial’ interests. Indeed, one way of characterising many of the assessment problems under discussion is that they arise from the substitution of private/partial valuations for public valuations (where by a public valuation we mean a valuation that reflects the preferences of the general public, in the manner, say, of a price determined in a competitive marketplace).

⁵⁶ This dichotomy is, quite clearly, a simplification, since hypotheses are not predestined to come either as a single spies or in battalions: there is a spectrum of possibilities.

Whichever type of hypothesis we are dealing with – whether it be an input into or an output from an economic theory or model – a single-hypothesis approach will focus on assessing the likelihood or probability that the particular hypothesis is true (or false). For a hypothesis H_0 let this be $P(H_0 | E)$, on the basis of an existing set of evidence or information E . This is referred to as a conditional probability – the probability/likelihood is contingent on the information that is available – and the symbolism can be read as “the assessed or perceived probability that hypothesis H_0 is true given, or contingent on, the set of information E .” (There is an question of how, appropriately, the assessor might ‘bootstrap’ herself/himself to initial likelihood, but this is a question that is more easily addressed after considering how assessed likelihoods might change in the face of new information/evidence.)

At least if it is relevant to the hypothesis of interest, new incremental evidence, ΔE , will change the assessed probability to $P(H_0 | E \& \Delta E)$, where ‘&’ indicates the combination/union of the two sets of information (i.e. the combination of the old and new evidence). One measure, or metric, for the value of the new information is the absolute value of the change in the assessed likelihood of the proposition of interest $ABS\{\Delta P\} = ABS\{P(H | E \& \Delta E) - P(H | E)\}$. The change in the probability reflects what can be learned from the incremental information, ΔE , and it is the absolute value of the change that matters, since the impartial assessor will not be concerned with whether the effect of the evidence is to increase the likelihood of the proposition ($\Delta P > 0$) or to decrease it ($\Delta P < 0$).

Going one step further, an investigatory process might set out to seek evidence/information that has most effect on the likelihood of interest; and such an approach would not be too far removed from some of the ideas of Popper. Whilst it appears inappropriate in the relevant, assumed context to adopt a falsificationist approach to an assumed hypothesis (implied hypotheses are a different matter) – e.g. deliberately setting out to find evidence that would contradict a proposition such as *the market price-elasticity of demand is constant* or *this undertaking was pricing below its average variable cost*⁵⁷ – Popper, in response to criticisms of his general position, also suggested a weaker version of that position, based on what he called the *principle of severe testing*.

This involves examining the likelihoods of observing certain things in the event that a hypothesis of interest is true, and in the event that it is not true (or possibly simply not on the decision maker’s radar). Starting from a general state of knowledge K , which resembles our starting evidence E but excludes any knowledge about the hypothesis of interest, then the expression $P(\Delta E | H_0 \& K) / P(\Delta E | K)$ represents the ratio the probability of observing a particular piece of evidence ΔE if the hypothesis is correct, compared with the assessed

⁵⁷ For example, an investigation may have been triggered by a complaint from a competitor alleging that it has been harmed by predatory conduct. In such circumstances, a falsificationist strategy might be held to be partial because it would seek to falsify the allegation of the complainant, rather than taking a more neutral approach to the possibilities. Similarly, a finding that the market price-elasticity of demand is not constant may not advance matters very far in a particular case, since it still leaves open a very wide range of alternative possibilities.

probability of that evidence on the basis of background knowledge only.⁵⁸ The larger the ratio of the probabilities, the greater the severity of the test; and Popper's recommendation is that the assessor should seek out severe tests.

This might be made clearer by a simple example. At the beginning of the twentieth century background knowledge in physics might be said to have attached high likelihood to the proposition that light travels in straight lines, and hence a low probability to any prospect of observing deviations of light caused by gravitation (the ΔE of this example). In contrast, the addition of the theory of general relativity to this background knowledge implied that light would be 'bent' by gravitational forces: $P(\Delta E \mid H_0 \& K)$ is a high probability. Eddington's 1919 testing of the theory, at a moment of solar eclipse, by observing whether light travelling from stars on the 'far side' of the sun from earth was bent or deflected by the sun's gravitational field could therefore be said to be a severe test of the theory, and hence a test capable of producing high value information.

The example is useful because Popper's severe testing approach has often been classified as a confirmationist approach to the evidence, on the ground it recommends seeking out evidence that has the biggest potential impact in confirming the evidence. The Eddington experiment indicates, however, that this approach may, quite obviously, lead to evidence that serves to falsify the theory: had the measurements been otherwise, something that was predicted by the theory of general relativity would have been observed not to happen, which would have been strong evidence against the theory.

The key point, therefore, appears not to hinge on a contrast between confirmationist and falsificationist approaches to evidence, but rather on seeking evidence that has a high impact on the perceived likelihoods of competing hypotheses. Such evidence can be said to be *highly diagnostic* in relation to the question(s) at hand, meaning that it is capable of conveying substantial new information about what is, or what has been, the case. In cases where the evidence is considered sufficient to reject a particular hypothesis outright, it can be said to be *decisive*.

The relativity example also serves as a reminder that the questions of interest will often involve a live interest in the assessment of more than one hypothesis or proposition: in the general relativity case there was the small matter of the competing predictions of classical theory on the one hand and of Einstein's new analysis on the other. In the limit, the alternative to the hypothesis that a proposition is true need be no more than its complement, that the proposition is false; and, if the complement is preferred on the basis of the evidence, there may be little or no significant implication for the perceived likelihood of any single alternative. But in economic assessments, as in the general relativity example, there may be 'live' alternatives, each attracting the allocation of a significant likelihood, such that those likelihoods would be liable to significant adjustments in the event of the discovery of

⁵⁸ Severity can also be defined in terms of the difference between the two likelihoods, rather than the ratio, but the basic idea remains the same.

evidence that was highly diagnostic in relation to any one of them. Roughly, if one proposition is discredited by a particular piece of evidence, the likelihoods of a number of other, well defined alternatives could be expected to be materially increased. In a sense, this is a form of competition among alternative views of the world that exhibits some analogous features to oligopolistic competition in economics (whereas, when the focus is exclusively on a single hypothesis, this competition from identified alternatives drifts into the background).

5.2 *The Bayesian approach*

In circumstances where the task is to evaluate the relative merits of two or more competing hypotheses, an obvious conceptual framework to adopt is that of Bayesian induction/learning. Starting from *prior* beliefs about the hypotheses/propositions, represented by probabilities or relative likelihoods, this provides rules for adjusting those beliefs in the light of new evidence. To illustrate, consider propositions H_1 and H_2 , to which are attached prior beliefs/likelihoods $P(H_1|E)$ and $P(H_2|E)$ on the basis of an existing evidence base E , and suppose that new evidence is available which has bearing on these beliefs. The question is: what effect should the new evidence have on the relevant *balance of probabilities*, as reflected in the ratio $P(H_1|\Delta E \& E)/P(H_2|\Delta E \& E)$, where $P(H_i|\Delta E \& E)$ is what, in Bayesian inference, is called the *posterior* probability – the revised assessment of the status of proposition H_i , once account has been taken of the new evidence.

By Bayes theorem:

$$\frac{P(H_1|E \& \Delta E)}{P(H_2|E \& \Delta E)} = \frac{P(\Delta E \& E|H_1)}{P(\Delta E \& E|H_2)} \frac{P(H_1|E)}{P(H_2|E)}$$

That is, the revised balance of probabilities is simply the *prior* balance of probabilities multiplied by the *likelihood ratio* (the first term on the right hand side of the equation), measuring the relative, assessed likelihoods of observing a particular piece of evidence if the relevant propositions are true. Put another way, the likelihood ratio $L(\Delta E \& E, H_1, H_2) = P(\Delta E \& E | H_1) / P(\Delta E \& E | H_2)$ determines the ratio of the adjusted balance of probabilities, which takes account of the new information ΔE , to the prior balance of probabilities. If $L(\cdot) = 1$ the new evidence does not give cause for any revision of views about the competing hypotheses, and is therefore not diagnostic; but the greater the deviation of the ratio from one the greater the diagnosticity.⁵⁹

The advantage of putting things this way is that, when an investigatory process is triggered by a particular proposition, or set of propositions, it helps keep to the fore the point that information/evidence about alternative propositions which may not itself be very directly related to the main proposition of interest can nevertheless be highly diagnostic. Consider,

⁵⁹ We are concerned here with general approach rather than with detailed measurement of the value of information, but it can be noted that an obvious measure of diagnosticity, which is symmetric in relation to the probabilities is $ABS\{\ln(L(\cdot))\}$, the absolute value of the natural logarithm of the ratio.

for example, evidence about which a hypothesis H_1 has little to say. If the likelihood of the evidence on an alternative hypothesis is, say, very high, then observation of that evidence can support a shift in the balance of probabilities toward the alternative. Put more simply, if one hypothesis does *not* explain some significant aspect of the relevant market context, whereas an alternative hypothesis does, other things being equal that should count as diagnostic evidence that tilts the balance of probabilities toward the alternative.

More generally, the likelihood ratio points to the importance of *differences in likelihoods* in the revision of beliefs in the face of new evidence, i.e. to the *diagnosticity* of the relevant evidence. What the assessor should be looking for is evidence that has significantly *different* likelihoods under the alternative, competing hypotheses. Thus, evidence with very high likelihood according to hypothesis H_1 will be of little diagnostic value if its likelihood is correspondingly high according to the alternative propositions of interest.⁶⁰ Similarly, H_1 may imply that some observed evidence has a low likelihood, but, if the alternatives imply that the evidence has a virtually zero likelihood, the observations can still be diagnostic; just as if H_1 implies a significantly higher likelihood.

This last point resonates with Sherlock Holmes's approach: *'It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth.'* In practice, it is generally impossible to exclude all alternatives but one; both because the evidence is rarely clear enough, and because it is generally not feasible for the assessor, acting under resource constraints, to consider all possible alternative explanations of things. Nevertheless, in a situation where there is a relatively small number of live alternatives, this type of 'eliminationist' strategy may be feasible.

It can also be noted that, in circumstances where the purpose of economic assessment is to lead to an adjudication, it may be that what is ultimately of interest is not an evaluation of the merits of individual hypotheses, but of groups of hypotheses sharing a single characteristic. In competition law, for example, the relevant question may be *did/does this type of business conduct, in this economic context, cause harm to competition and to consumers?* There may in a particular case be many theories that imply that the answer should be yes, and many that imply that it should be no, and it is evidence that is diagnostic as between the two groupings that is most relevant.

In this case, the diagnostics potentially become much simpler, at least when *ex post* assessment is feasible: rather than focusing on testing a single theory of harm or implied hypothesis – which, after a good deal of effort, may, like the vast majority of economic hypotheses, be found to be false, hence advancing knowledge only marginally – the direct, empirical question can be asked: *is there evidence linking actual harm to the conduct in question in the relevant context?* It is the potential diagnosticity of such evidence for the

⁶⁰ A classic example in competition law is price parallelism in a market for a homogeneous product. Both competition and cartelisation tend to imply a degree of parallelism, so that an observation that price movements are highly correlated has little diagnostic value by and of itself.

issue of interest (economic harm) that underpins the arguments for an effects-based approach to competition law enforcement.

5.3 Evaluating multiple propositions and the problem of the scope of relevant information

Even when the central interest is in the validity or non-validity of just one, particular theory or implied hypothesis and its variants, where feasible it is usually helpful in assessment to keep several competing hypotheses in mind. As discussed above, theories tend to be built up from sets of assumptions/propositions, and changing one or more of the assumptions may lead to different implications. Whether such changes are said to lead to different theories or to variants of a single theory is a matter of semantics: what matters is the variation and the implications of the variation.

Because the potential variations are many, it is generally infeasible to evaluate all possibilities, and it is here that the significance of context becomes more apparent. Careful study of context, starting with background fact-finding, is a way of narrowing the field of live, competing hypotheses. More specifically, many of the simpler ‘assumed hypotheses’ (or, more simply, the assumptions) to be relied upon can be settled on the basis of an appreciation of the background or contextual factual material.

Since these ‘joint propositions/hypotheses’ tend to contain large numbers of elements, the usual approach is to seek to fix as many of the elements as possible by reference to factual investigation of the relevant context. Economic assessment is not, therefore, typically a process in which some well-defined joint-hypothesis is first developed and only later tested against the facts. Rather it starts (or should start) with fact finding, since fact finding is generally the only effective way of developing a joint set of hypotheses in the first place, at least without straying too far from market realities. The question then is: how much information needs to be collected?

Although there are no simple answers to this question, there does appear to be significant evidence that, in practice, the search tends to be too narrow, and that, both individual decision makers and organisations, tend to ‘jump to conclusions’ on the basis of highly limited information sets and an inadequate appreciation of the economic context (the *premature closure* problem). This may be linked to a systematic tendency of fast, inductive thinking to seek for coherent accounts of events and of the ‘state of the world’. Such a demand is easier to satisfy when the range of facts to be explained is relatively limited: limited factual evidence gives more freedom in the use of auxiliary hypotheses, assumptions and propositions in developing a theory or narrative. Simple stories are easier to tell: they leave many more things ‘unexamined’ and are less constrained by facts. In everyday language, they leave more things to the imagination, which is a positive for storytelling, though not for accurate assessment and adjudication.

This is, perhaps, the major source of confirmation bias in economic assessment for public policy purposes: a simple story becomes privileged at an early stage of the process. The

information discovery process then tends to falter or come to a halt; not necessarily as a result of a conscious decision, and possibly in consequence of the fact that, once a storyline has emerged, new information becomes a source of potential incoherence.⁶¹

The development of multiple, explicit theories/narratives/stories at an early stage of the assessment process can potentially mitigate some of these problems, even if all of them are false. Different theories tend to have different implications, which can point towards the types of evidence that may be *diagnostic*. And evidence that is highly diagnostic as between any two sets of propositions (i.e. theories) will typically shift the likelihoods of other theories as well. Thus, a simple story that initially appeared to have coherence and explanatory power in the context of early evidence, may come to be seen in a different light when information gathering that may have been triggered by some other lines of reasoning has been completed.

Competing propositions, hypotheses and theories tend to assist assessment by establishing multiple perspectives from which both existing evidence/information and the search for new evidence/information can be viewed. This will tend to increase the chances that evidence/information with significant diagnostic power will not be missed, discounted or rejected; and there are obvious analogies here with the discovery properties of competitive markets, which also rely upon variety among the perspectives of market participants.

The underlying, inductive logic is no different in the case of multiple hypotheses as it is when there is a single hypothesis/theory of interest. Where the focus of attention is on a specific theory – say a theory of harm in a competition policy case – the relevant balance of probabilities becomes the ratio of the probability that the theory of interest is true to the probability that it is false. A lower bound for the latter is the sum of the probabilities attaching to the truth of the competing theories (it is only a lower bound because other possibilities may exist which are not considered at all – the ‘black swans’ whose relevance should always be recognised, but whose quantification is always likely to be problematic). The merit of developing explicit alternatives, however – as distinct from a generalised ‘not so’ alternative to – is that it requires the explicit development of alternative perspectives on the issues, which is generally good for learning and discovery. In technical jargon, it can be said to promote *integrative complexity*.⁶²

⁶¹ Organisations may possibly have a greater propensity than individuals wilfully to reject and discount ‘disturbing’ evidence, because they have problems of social coherence and integration to address. The normal preference for coherence in narrative is, therefore, strengthened by the functions that the relevant narrative serves in maintaining organisational coherence. In relation to public policy decisions, Barbara Tuchman has put the point in rather strong terms: ‘Once a policy has been adopted and implemented, all subsequent activity becomes an effort to justify it’, *op cit*. In our experience, this process of ‘justification’ begins rather earlier (than the time of policy adoption), and it has its genesis at the moment at which an early, coherent storyline emerges (although its full force does not emerge until later).

⁶² Integrative complexity is a psychological metric that refers to the extent to which individuals demonstrate two inclinations when they consider events and issues. The first dimension, differentiation, relates to the capacity of individuals to adopt and to apply a variety of perspectives to appreciate an issue. The second dimension, integration, refers to the capacity of individuals to recognize connections and similarities across divergent

5.4 Starting points and counterfactuals

As indicated above, the Bayesian approach to inductive logic requires a starting point in the form of initial or *prior* probabilities or likelihoods. Where a particular hypothesis or theory is the centre of attention, it is sometimes argued that there is a case for a 50-50 starting point for a balance of probabilities assessment – a starting point that will then be modified by assessment of evidence. We have already indicated one reason why we believe that such an approach is likely to promote error – theories are combinations of propositions, and fact finding and an appreciation of context is required at the outset, before there can be any realistic question of attaching likelihoods to competing alternatives. The development of multiple competing theories/propositions provides further protection against assigning an over-inflated, initial likelihood to a privileged theory. The alternatives serve as a reminder that the initial ‘story’ may be one of many stories that are coherent with the limited information that might be available at the beginning of an assessment process, and that most simple, coherent stories are likely to be false.

Making use of multiple, competing, explanatory narratives can also assist in avoiding what seems to us to be becoming a systematic error in the enforcement of competition policy. This error occurs when a theory of harm – and there is usually only one theory of harm to be considered – is evaluated by comparing actual or expected market outcomes with a *single* counterfactual position.

The logic here is clear enough: in a case of alleged abuse of dominance, for example, the idea is to examine what might have been expected to happen ‘but for’ the business conduct of interest. The difficulty is equally clear – a number of different things might have happened – and it is related to the obstacles to performing controlled experiments in the kinds of complex, adaptive economic systems and sub-systems with which public policy is concerned. Too much is ‘going on’ to have any realistic hope of being able to forecast precisely how things would have evolved, or (in the case of *ex ante* evaluations) will evolve ‘but for’ the business conduct in question; and to assume otherwise is to fall into the fundamental (and persistent) error of pretending that the future is significantly less uncertain than it actually is.

Realistic ‘effects analysis’ addresses the problem of uncertainty, rather than avoiding it, and one of the ways it does this is by explicit acknowledgment of the multiplicity of stories that can potentially be told, and by explicit examination and comparison of those stories in the light of available evidence and of evidence/information that could potentially be collected and that might also be capable of casting light on the validity of the competing alternatives. Among other things, this means trying to look at issues from alternative perspectives, and in different ways, and on the basis of different parts of the wider evidence and information base.

perspectives. Hence, when integrative complexity is low, individuals tend to form simple and rigid attitudes and perceptions. A similar notion can be applied to social groups, including assessment teams within organisations.

5.5 Cumulative evidence, contra-indicators and the zero-one bias

Diagnostic evidence is information that materially affects the relative likelihoods of competing, alternative hypotheses and theories being valid. In the course of an economic assessment, it may turn out to be the case that some evidence significantly reduces the relative likelihood of a particular hypothesis being true, whilst other evidence has the reverse effect. That is, individual pieces of evidence, whilst diagnostic, do not, when considered as a whole, lead to any very radical shift in an initial balance of probabilities.

In other cases, individual pieces of evidence may each have similar effects on the likelihood of a particular hypothesis, whether by systematically increasing or by reducing the likelihood. In these circumstances, it can be said that the evidence tends to ‘stack up’. At the end of the assessment process, the perceived likelihood of a particular hypothesis may have changed very substantially, even though no one piece of evidence has, by and of itself, led to a dramatic shift in the balance of probabilities. That is, *individual* pieces of evidence may not be *decisive*, yet their cumulative weight may be.

It is, then, the cumulative implications of a range of pieces of evidence, which may have been obtained from disparate sources and which may have cast light on disparate aspects of the issues under examination, that ultimately matter. Unfortunately, economic reasoning in competition law cases indicates that, in all too many cases, individual pieces of evidence are assessed on the criterion of *decisiveness* rather than *diagnosticity*. The error appears to be pervasive, and we will give examples in the next section.

This systematic error, or bias, is frequently just another aspect of the tendency toward confirmation bias. If a privileged theory or hypothesis has been developed, there may be a number of potential *contra-indicators* to that theory which have been brought to the attention of the assessors (by a contra-indicator we mean a piece of diagnostic information that might be expected to reduce the likelihood of the theory being true). A common approach is to examine a piece of relevant information in isolation from other factors, conclude that the evidence/information is not sufficient to cast serious doubt on the maintained theory (for example, because the evidence is capable of being explained in ways that are consistent with the theory⁶³), and then, crucially, to set it aside and not consider it further.

It is the last aspect of evaluation practice that is the problem here. It strips the individual piece of evidence from the overall context and isolates it, making it easier to dismiss it. It is not unusual to find competition law decisions repeating this exercise for a list of three, four, five, or more contra-indicators, without any sign of awareness that, taken in its entirety, the available, cited evidence comprehensively undermines the favoured theory.

⁶³ At this point, there is often some tacit adjustment of auxiliary hypotheses/propositions in a theory, which amounts, in effect, to a change in the joint hypothesis being tested, although such adjustment is rarely made explicit.

The effect of this ‘reductionist’ approach to the evidence base is that evidence and information that is (a) diagnostic/relevant, but (b) not decisive by and of itself, is excluded from consideration: there is a *zero-one* or *all-or-nothing* approach to individual pieces of evidence. Inconvenient evidence can thereby be evaded, and simple storytelling can be bolstered. If a piece of evidence fits the story, it is included; if it doesn’t fit the story, it is neglected.

It is therefore somewhat ironic that persistent perpetrators of the error are sometimes also advocates of ‘economic’ approaches to competition law enforcement, who emphasise the importance of economic analysis in the face of the complexity of the market contexts with which the authorities have to deal. Such advocates are right about the underlying complexity, but the errors and biases that currently pervade economic assessments, at least in the European jurisdictions with which we are familiar, indicate that, thus far at least, often all that has happened is that one form of simplistic thinking has been replaced by another.

6. Specific examples of diagnostic failures in competition policy.

In an attempt to give flesh to some of the points made above, we now give two examples, based on our own experience, of competition assessments where confirmation bias appears to have been at work, and where there have been failures to recognise or to seek out relevant, diagnostic evidence. We also give an example of a case in which the diagnosticity of particular evidence was appropriately recognised in the conduct of a competition law investigation, to indicate that good practice is perfectly feasible, and not just a utopian dream.

6.1 *The Monopolies and Mergers investigation of beer supply: 1986-9.*

The MMC’s now relatively ancient investigation of the UK beer market was concerned chiefly with the effects of vertical integration between the brewing and retailing of beer, and the case continues to have relevance in illuminating attitudes to the evidence that have been repeatedly signalled in cases since. The investigation ran for two and a half years and involved the processing of very large amounts of information, particularly accounting information. However, it has been said since, including by some of those involved, that minds were made up very early in the process, and the subsequent progress of the investigation and reasoning in the final report seem to us to be consistent with this view.

The particular part of the evidence to be highlighted here concerns international comparisons of industry performance. It was a fact that vertical integration of brewing and retailing was particularly pronounced in the UK, matched only among the beer major markets by the Republic of Ireland. The Brewers Society therefore argued that international comparisons of performance were a potentially good way of assessing the effects of vertical integration: that is, in the terminology we are using, such studies were potentially highly diagnostic. The

Society therefore conducted international comparative studies of pricing in retail on-licences and of consumer views of how British pubs compared with their nearest equivalents abroad.

Table 1 below summarises the principal results of one of the exercises, which involved an international comparison of the prices of major beer brands in on-licence premises (where vertical integration was very high in the UK by international standards) with prices in off-licence premises (where vertical integration is low in the UK, as well as elsewhere, and where international trade in beers sold in supermarkets could be expected to have equalising effects on net-of-tax prices among countries). Column 1 shows the average ratio of on-licence to off-licence selling prices for beer in the sample countries, while column 2 shows the net-of-duty price ratios. The Table indicates that the *relative* price of beer in UK on-licensed premises was substantially lower than in continental Europe and significantly lower than in the USA.

TABLE 1 Beer price ratios, 1987

Country Ratio between on on-licence and off-licence prices

	<i>With tax</i>	<i>Without tax</i>
Belgium	3.1	3.2
France	3.5	3.6
Italy	3.8	3.8
USA	2.3	2.4
West Germany	3.5	3.6
United Kingdom	1.7	2.0

Source: The Supply of Beer, MMC Report, 1989

As always with international studies, and indeed with more or less any comparative studies, this type of work does not, and can not feasibly correct, for all contextual differences; which is simply to repeat again that, in economics, we are rarely able to do controlled experiments. From at least the time of Hume, however, it has been recognised that, notwithstanding the imperfections, international comparisons can contain diagnostic information.

The questions raised by the MMC in relation to the studies related to a number of factors: the relatively small size of the sample (12 – 30 outlets of each type in each country); differences in product mixes across countries (particularly the higher proportion of draught beer sold in UK on-licensed premises, which has lower packaging costs per unit); differences in the services provided alongside beer within an on-licensed outlet (although that aspect was partially dealt with by a parallel study of consumer views of the ‘amenity’ aspects of on-licensed premises); economy-wide variations in the relative prices of services to manufactures across countries (associated with differential productivities); and differences in

the relative significance of beer in the overall offering of on-licensed premises across countries.

These were reasonable questions to raise, and the MMC was undoubtedly correct to point out that *“All of these factors may contribute to the difference in the ratios between countries and it cannot be concluded that the difference is due to vertical integration in the United Kingdom (even if all the other countries were free of vertical integration, which they are not).”*

But what is striking about the comparisons are the sizes of the differences: the net-of-duty relative price of beer in on-licensed premises in West Germany was around 80% higher than in the UK, for example. The cumulative effect of adjustments to reflect other factors that might be expected to have influence on the relative price – including those factors that might have worked in the opposite direction to those listed by the MMC (e.g. property prices) – would have needed to have been very substantial to put UK performance in a poor light relative to the comparator countries. And, if the MMC had suspected that was the case, it would have been relatively straightforward to assess the relative quantitative significance of the factors of interest, by undertaking further work in this area – all of the factors mentioned were amendable to quantification – which could have been done using only a fraction of the resources actually allocated to accounting studies during the course of the investigation.

The fact is that the MMC panel in this investigation was not interested in this line of inquiry, and (with the distinguished exception of the author of the minority report) was content to conclude that: *“... in view of our substantial reservations about the survey we do not think that it would be safe to conclude that on-licensed beer prices are good value for money compared to those in other countries.”*

This statement gives the game away: there is manifest assessment error at this point, and it has at least three aspects that are worth noting explicitly:

- *The wrong question is asked.* The MMC has assessed the evidence on the criterion of its *decisiveness* for the hypothesis of interest: “on-licensed beer prices are good value for money compared with those in other countries”. The more important questions should have been: Is the evidence *diagnostic* for the question at hand? Does it lead to a non-trivial adjustment in the balance of probabilities concerning the likelihood of alternative conclusions being valid, and might it continue to do so when appropriate account has been taken of other factors at work?
- Having concluded that the evidence is not decisive (a right answer to a wrong question), *the evidence subsequently received a zero or near zero weight in the MMC’s deliberations.* This is an example of the zero-one or all-or-nothing tendency mentioned above. Further, since the relevant evidence was plainly diagnostic or

potentially diagnostic, its dismissal amounts to a failure to properly consider relevant evidence – notwithstanding the explicit discussion of the evidence, and the explicit rejection of it.

- Given that the international comparison evidence was potentially highly diagnostic – for example, because (a) the off-licence prices for beer provided for a benchmark that standardised for a number of economic factors that would otherwise have complicated international price comparisons and (b) the degree of vertical integration in the UK on-licensed supply of beer was very substantially (not just a little) higher than was typical in comparator countries – the MMC should have pursued matters further, to try to increase the value of the relevant information. Instead, *no attempt was made to collect additional information on the pricing pattern.*

In our experience, these types of errors occur time and time again.⁶⁴

6.2 *The Intel case*

Two decades after the Beer investigation in the UK, we find a very similar approach to inconvenient, diagnostic evidence in the European Commission’s Decision in *Intel*.⁶⁵ The case concerns the economic effects of discounts on supplies of central processing units (CPUs) offered to original equipment manufacturers (OEMs) such as Hewlett Packard, Dell, Acer, Toshiba, Sony and others. The question of interest to the European Commission was whether the discounts were an abuse of dominance on account of having significant foreclosing effects on the market.

Since supply arrangements in the relevant market(s), including pricing, tended not to be based upon detailed, formal contracts, but rather, as the Commission found in its fact finding, on informal bargains and ‘handshakes’, this was not a case in which OEMs could read off from pricing schedules how much discount they were liable to lose in the event that they switched business away from Intel (it was different in this respect, therefore, from older conditional discounting cases such as *Hoffman La Roche*). Hence, the Commission’s case was necessarily based upon propositions concerning the *expectations* of OEMs about the consequences of such switching of business. As the Decision explains, the Commission’s view (at para 306) was exemplified by its interpretation of the supply arrangements between Intel and Dell:

⁶⁴ The frequency of the problems can also be inferred from, for example, the references to the ‘in or out’ tendency in the work of Franklin Fisher on market definition. It has also been referred to as the ‘zero-one fallacy’ in the market definition context, by Sir John Vickers. See FM Fisher, “Horizontal Mergers: Triage and Treatment”, *The Journal of Economic Perspectives*, Vol. 1, No. 2 (Autumn, 1987), page 27.

⁶⁵ Commission Decision relating to a proceeding under Article 82 of the EC Treaty and Article 54 of the EEA Agreement (COMP/C-3 /37.990 - Intel) 13 May 2009.

“... the Commission has not stated that Intel had an agreement with Dell which precluded Dell from purchasing AMD microprocessors, as it eventually did in 2006. The Commission's findings, as illustrated in sections 2.3.4.1 and 2.3.4.2, are that Dell was free to start sourcing x86 CPUs also from AMD, but that this would have entailed the loss of a significant and disproportionate part of the Intel MCP rebates. In other terms, Dell always retained the right to choose AMD, but this right to choose was exerted in a context where opting for AMD would have had a disproportionately negative impact on the rebates that Dell obtained from Intel (our emphasis).”

Later, in explaining the incentives not to switch business (at para 939), the Decision makes clear the significance for its reasoning of the beliefs/expectations of Dell:

“This economic inducement mechanism functions irrespective of whether the benefits to the customer of staying exclusive are retrospective or prospective, and irrespective of whether they are guaranteed by contract or tacitly agreed. The mechanism works as soon as there is a sufficiently clear understanding for the customer that the dominant company will award it certain rebates if it remains exclusive, and will not award them if it buys from competition (our emphasis).”

The evidence in the case regarding Dell was of particular significance since, during the relevant period, it was (a) the only OEM that relied exclusively on Intel CPUs in its computers, and (b) the largest of the OEMs in the early part of the relevant period, although it was later overtaken by Hewlett Packard (which purchased substantial volumes of AMD CPUs). The question of whether Dell was deterred from switching some of its business to AMD (which it eventually did) by expectations of disproportionate reductions in discounts, or chose not to switch for legitimate commercial reasons (e.g. universal co-branding of its machines, using the Intel brand name as well as its own), was therefore of considerable importance.

Clearly, a central issue here is what is meant by the notion of disproportionate loss of discounts. In the normal course of events, it is neither unusual nor problematic for suppliers to offer larger per unit discounts to customers who purchase higher volumes (i.e. for suppliers to practice volumetric pricing). This, however, is a familiar problem; and we focus here on the more novel aspect of the Commission's reasoning, which, given the lack of formal contracts and price schedules, concerns assessing whether there was formation of a 'clear understanding' that a disproportionate loss of discounts could be expected in the event of significant switching of business away from Intel.

During the course of the investigation, internal Dell emails were produced that indicated no such clear understanding existed in the company. The Decision makes reference to one of those emails at para 266,

" – *there are two schools of thoughts within Dell. One side believes that Intel will hurt us and hurt us bad and the other side is that things will get better even if they hurt us in the short term.*"

and assesses this evidence as follows:

“Contrary to what Intel argues, this document does not show that there was a way of thinking within Dell (that is, a "school of thought") according to which introducing an AMD product would not result in any Intel reaction. On the contrary, the above document makes clear that the unanimous belief within Dell was that switching, in part, to AMD would result in Intel "hurting" Dell. The discussion in the email within Dell then focused on whether in the long term things would improve despite Intel hurting Dell in the short term.”

Once again, there are errors of assessment here. First, and as in the beer case discussed above, there is a shift in the question asked, from the relevant question “did Dell believe that there would be a *disproportionate* reduction in discounts if it switched business to AMD”, to an irrelevant question “did Dell believe there would be an Intel response to business switching?”

Second, the e-mail simply does not indicate a unanimous belief within Dell that switching would result in harm to Dell. The Commission’s interpretation might be defensible if the second of the quoted sentences from the email had said “... things will get better although they will hurt us in the short term”. But it doesn’t say that: it says “*things will get better even if they hurt us in the short run.*” That is, it is a *conditional statement*, not a statement about what is expected to happen. The effect of the use of ‘even if’ is actually to strengthen the argument of the school of thought that appears to have taken the view that it would be beneficial to switch some business from Intel. It amounts to a familiar rhetorical device whereby it is argued that, even in a worse-case scenario (being hurt in the short run), a switching policy would still be advantageous. By implication, switching business would be significantly more advantageous still if the worse-case scenario (being hurt in the short run) did not eventuate.

The underlying economics here is that commercial buyers will often face situations where, other things equal, they might expect some loss of volumetric discounts if they switched some business away from their main supplier; but at the same time they might expect some positive commercial response from the supplier consequent on the tightening of the competitive screws to which switching might give rise (i.e. the supplier will be chastened by the loss, and increase its efforts to provide better value for money to the customer).

The important point, however, is not so much that there is a mistake in the Commission’s logical reasoning, but that the mistake correlates with a failure to explore the possibility of differences of views within Dell about its single sourcing policy. That is, the fallacious reasoning has the effect of providing a specious justification for dismissing the email, not investigating further, and therefore not addressing the issues and questions that the email

should properly have raised: What views existed within Dell? Who held them, and what influence did they wield? On what were the views based? and so on. All of these questions are relevant to assessing the proposition that there was a “*clear understanding*” on the part of Dell (and other OEMs) that Intel would respond to loss of business in ways that would have “*a disproportionately negative impact on the rebates that Dell obtained from Intel*”, rather than, say, by reducing discounts in ways that did no more than reflect increases in unit costs of supply associated with lower volumes, or by increasing its efforts to offer better deals to Dell to win back lost business.

The obvious inference is that the Commission chose not to seek out further evidence because the result of such an exercise could be expected to be an increase in the likelihood of finding facts that would be contrary to an established economic narrative favoured by the Commission’s assessors.

6.3 White rum

Lest the discussions thus far might suggest the thought that what we are arguing for is itself an aspiration beyond possible attainment, and that current levels of performance in economic assessments will always be with us, our final example is one in which a public authority, in this case the UK Office of Fair Trading (OFT), *did* respond to the discovery of new, diagnostic evidence in ways closer to those of an idealised adjudicator or skilled, engaged impartial spectator. It concerns market definition in relation to the supply of white spirits, with particular reference to white rum.

Briefly, the OFT was developing a Competition Act case against Bacardi for abuse of dominance in a market defined as the supply of white rum in on-licensed premises. The investigation had involved an assessment of whether white rum was sufficiently substitutable with other white spirits – vodka and gin – to constitute part of a wider market, and the assessment had been based in part on an econometric study that had found low cross-price elasticities of demand among the different white spirits (and hence supported a preliminary conclusion of separate product markets).

In the course of the investigation, however, a remarkable and relevant fact was discovered from marketing evidence: nearly 90% of Bacardi sold was sold mixed with cola. That is, what the vast majority of customers were choosing between (in on-licensed premises) was rum-and-coke and vodka-and-coke. It should be immediately obvious that this is information that is potentially highly diagnostic. Putting it very simply, end consumers were buying alcoholic colas with moderately different flavourings. Or, put another way, the strong effects of the cola contribution to the mixed drink – and cola accounts for the bulk of the volume – homogenises the taste of the alcoholic components, and makes them more similar to the consumer. The end effect is that a much higher degree of substitutability in demand for the

alcoholic components might reasonably be expected than if it had been the case if end consumers drank the two spirits unmixed.

This evidence is not itself decisive for the market definition issue, but it opens up a new set of questions, leading to expansion of the available information set in ways that do, when the evidence is considered as a whole (i.e. cumulatively), lead to fairly clear cut results. For example, the consumption pattern explains why, in the recent past, Bacardi had re-engineered its product to suppress some of the stronger and more distinctive rum flavours from its product. Normally one might have expected that, for leading brands of white rum and vodka such as Bacardi and Smirnoff, that the suppliers would tend to differentiate their products in order to soften price competition. The likelihood of observing product homogenisation, on prior, general economic knowledge, would therefore tend to be assessed as relatively low.

On the other hand, if the aim is to win ‘alcoholic cola’ business, there is an obvious potential rationale for wanting to avoid an over-strong rum influence on the taste, and the assessed likelihood of such product engineering will tend to be significantly higher. In these circumstances, an observation of product re-engineering gives grounds, via the Bayes formula and the likelihood ratio, for a significant revision of the balance of probabilities toward a wider product market definition, encompassing (at least) both vodka and white rum.

As well as stimulating the information discovery process, the observation that most white rum was consumed together with cola also gave rise to a re-evaluation of the econometric evidence. It transpired that, on closer inspection, the relative prices of white rum and vodka supplied to on-licensed retailers had not materially changed over the sample period, and that the variations in prices that appeared in the data set were chiefly sampling effects: outlets were continuously sampled, but the set of outlets visited in one period was not identical to the set of outlets visited in the next period. Since on-licensed prices of alcoholic drinks can vary significantly, both by location and by precise type of outlet, these changes in sample composition led to changes in sample prices, even in circumstances where, in reality, the prices charged in each outlet remained unchanged. Volume figures were subject to similar sample composition effects, and it therefore became clear why, in the relevant factual context, significant correlations between recorded prices and volumes could be very low, even in the event of white rum and vodka being in the same market (i.e. even in the event of high substitutability between the two spirits). In terms of our terminology, the likelihood of observing low cross-price elasticities in an econometric study was very high irrespective of whether white rum and vodka were weak or strong substitutes, implying that the low estimated cross-elasticities constituted non-diagnostic evidence.

In the event, the OFT, did not proceed to a decision finding a breach of the Competition Act, and its response to the marketing evidence stands in contrast to that of the European Commission in *Intel*. In each case, a piece of evidence became available that contained a significant diagnostic signal and/or suggested a rich source of diagnostic evidence (the

existence of a documentary reference to two schools of thought within Dell; the fact that most white rum is consumed by mixing it with cola), but whereas the OFT recognised the evidence's significance and acted appropriately, the record suggests that the Commission recognised the relevant information's potential significance (hence the attempt to justify its dismissal) but chose not to let it disturb its prevailing beliefs.

7. Concluding comments

The discussion has covered a wide range of issues, and has been motivated by a desire to stimulate discourse on economic assessment rather than to provide definitive answers to questions raised. We will not attempt to summarise, but will simply emphasise four points in concluding.

The first is the importance of seeking *diagnostic* evidence. Theories, hypotheses and stories abound: it is *diagnostic* evidence that serves to narrow the field. The qualifying adjective (diagnostic) is important, because it is easy for calls for evidence-based approaches to degenerate into a search for evidence that confirms or justifies a favoured or privileged pre-chosen view of the world.

Second, the search for such evidence will, in and of itself, tend to broaden the information base upon which assessment proceeds. This should assist in counteracting the tendency of economic assessors toward over-abstraction in analysis, assist in the building up of a fuller picture of the relevant, factual context, and facilitate closer integration of theory with the evidence.

Third, for diagnostic evidence to distinguish between alternative hypotheses and theories, those hypotheses and theories need to be developed in the first place. There is a two-way interaction here, since hypotheses and theories are most effectively developed from available factual evidence, so as to offset the tendency of economic analysis toward over-abstraction. This creates a tension in that the early development of theories, to assist the search for diagnostic evidence, can increase the risks of *premature closure* (e.g. by rejection of an early-stage version of a hypothesis or theory, that would have been capable of useful development in the light of emerging evidence). What is required is *integrative complexity* in the assessment process, i.e. an ability to sustain, until quite late in the day, a number of alternative, 'competing' perspectives on the problems of interest, and to continue to use those different perspectives in assessing how the evidence as a whole 'stacks up'.

Finally, creating the organisational environments in which the search for diagnostic evidence can be pursued effectively is likely to be a difficult task. It is trite economics to say that information has value, but the value concerned will depend upon the criteria by which it is assessed, by the 'values' underpinning the assessment process itself. Information may, for example, be sought for economic gain, or to further the interests of a particular organisation or group. We have suggested that the relevant values are those of *skilled and engaged*,

impartial assessors, but those values are much easier to identify than to sustain in decision-making contexts.

Emphasis on *diagnosticity* is, in many ways, counter to the prevailing values of government policymaking: it implies, for example, that theories and narratives are disposable, whereas political and organisational life tends to be full of strong attachments to simplified narratives. Indeed, the diagnostic value of information derives from the fact that such evidence is capable of distinguishing between competing theories and explanations/accounts of the way things are. As with competitive processes more generally, some theories/explanations will win and others (likely to be much larger in number) lose. It places theories and narratives under constant risk of being undermined, and threatens wooden-headedness and ‘terrible simplifications’ in policymaking. As an approach to policymaking, therefore, it itself is likely to be under constant threat.

Fostering *communities and cultures of inquiry* within parts of government may perhaps be likened to creating protected habitats in an otherwise hostile environment. It is not something that can be done easily, that is likely to occur spontaneously, or that can be expected to lead to a dominant culture. Yet that licensed truth-telling in the courts of power is worthwhile is old wisdom, not new: “*Jesters do oft prove prophets*” (*King Lear*, Act V, Scene 3, Regan). Without it the risks of folly are greatly enhanced.