Fairness in Public-utility Regulation: A Theory

DARRYL BIGGAR

Abstract

Regulators routinely and systematically depart from policy prescriptions that are soundly based in conventional economic theory. In doing so, they often appeal to notions of fairness, justice, or reasonableness. Economists have historically struggled with these notions, which seem to be separate from, or in conflict with, conventional economic efficiency. This paper identifies five stylised facts about public attitudes to fairness in utility pricing, and argues that these stylised facts can be explained as an implicit attempt to protect the sunk investments of consumers in a natural monopoly’s services. Thus the paper suggests that the notion of fairness is not in conflict with the conventional notion of economic efficiency, but can be seen as consistent with the desire to promote sunk investment by the monopolist and its customers.

Introduction

Open any textbook on microeconomic theory or natural monopoly regulation and you will learn that the primary economic harm from natural monopoly is the allocative efficiency loss. First-year economics students are taught that an unregulated monopolist will produce ‘too little’ and set a price ‘too high’ relative to the theoretically efficient level. The ensuing reduction in economic welfare is known as the ‘deadweight loss’. According to the conventional theory, the primary objective of the economically literate and enlightened regulator is to regulate in such a way as to minimise or eliminate the deadweight loss.

1 Darryl Biggar is an economic consultant to the Australian Competition and Consumer Commission and the Australian Energy Regulator, Darryl.Biggar@accc.gov.au. The views expressed here are those of the author and are not necessarily the views of the ACCC or the AER. The paper has benefited from the comments of an anonymous referee.
During the twentieth century, a large body of economic theory was developed showing how a regulator should set tariffs so as to minimise the deadweight loss under different assumptions and constraints. This theory highlights the benefits of, amongst other things, marginal-cost pricing, various forms of price discrimination, Ramsey pricing, and peak-load pricing. However, although there has been some limited acceptance of these policies, the response of the regulatory community has, on the whole, been cautious and muted. Experience shows that regulators routinely depart from policies which are soundly based in conventional economic notions of efficiency.\footnote{2 See, for example, the citations in Biggar (2009) and later in this paper, such as Berg and Tschirhart (1995: 324), and Baumol (1986).}

Instead, regulators argue that these conventional economic pricing principles often do not satisfy their moral or legal obligation to take into account notions of ‘fairness’, ‘equity’, ‘justice’, or ‘reasonableness’. Many regulators, especially in the US, are statutorily obliged to set rates which are ‘just and reasonable’ or ‘not unduly discriminatory’. Even in the absence of an explicit legislative requirement that rates be ‘fair and reasonable’, regulators rarely adopt rates which — even though soundly based in conventional economics — might violate notions of fairness or equity.

Economists have historically reacted to the apparent trumping of efficiency concerns with notions of fairness or equity in a variety of ways. As discussed below, one approach has been to dismiss or belittle these other concerns, viewing them as illegitimate deviations from the pursuit of economic welfare. Another approach views fairness and equity concerns as potentially legitimate, but outside the professional sphere of the economist. There have also been several attempts to formalise and/or axiomatise notions of fairness and equity.

This paper argues that key elements of the notion of fairness in public-utility regulation can be explained through the lens of economic efficiency, using a different underlying economic model. Specifically, I argue that fairness concerns in public-utility regulation can, to a large extent, be explained as an explicit or implicit concern on the part of a well-intentioned regulator to protect and thereby promote sunk complementary investments by consumers.

Following the argument in Biggar (2009), I point out that users of a monopoly service must typically take some irreversible action which increases the value of the monopoly service — such as the decision of a large gas consumer to locate close to a gas-transmission pipeline, or the decision of a factory to install electrical wiring on its premises. The value of such investments is contingent on continuing to receive access to the monopoly service at reasonable prices and quality. Such an investment is therefore at risk of expropriation through an
increase in the price or a decrease in the quality of the monopoly service. The fear of such expropriation has a chilling effect on such investment, reducing overall economic welfare.

I suggest that regulators (in common with the rest of the population) are intuitively sensitive to the notion that their pricing policies could threaten the incentives of users to make valuable sunk investments. In rejecting certain pricing policies as ‘unfair’ or ‘inequitable’, well-intentioned regulators implicitly have in mind the impact of those pricing policies on the incentives on users to make sunk complementary investments.

This hypothesis seems to explain the stylised facts about attitudes to fairness in pricing, set out below. Furthermore, this proposition is, at least in principle, empirically testable. For example, I suggest that fairness concerns are important precisely in those sectors in which the value of the complementary investments by consumers are dependent on the price and quality of the service they receive from a specific firm. Conversely, I suggest that fairness concerns are, in practice, diminished in those industries where customers’ complementary investments are not specific to any one firm — that is, those industries in which customers face effective competition.

First, however, section 2 makes the case that there is a need for economists to study and understand fairness. This section also distinguishes notions of fairness which are not directly relevant to this paper and briefly addresses the concern that perhaps fairness is not systematic enough to be a legitimate subject for study. Section 3 surveys the responses of economists to fairness concerns in the past. Section 4 derives some stylised facts on attitudes to fairness in pricing. Section 5 shows how those stylised facts are consistent with a concern to protect sunk complementary investments.

What is fairness and why is there a need to explain it?

Economists — especially those at the ‘coalface’ of interaction with the regulatory community — have long been aware that regulators do not always do what conventional economic theory tells them they should do. Even the most well-established pricing policies in conventional regulatory economics are often downplayed or overruled by regulators.

For example, economists have long observed that Ramsey pricing and peak-load pricing, although solidly based in conventional economic theory, have been only partially accepted by regulators. In rejecting these policies, regulators routinely
point to concerns about fairness or equity (see, for example, Kahn 1970; Baumol 1986; Faulhaber and Baumol 1988; Berg and Tschirhart 1995; Raux and Souche 2006).  

This focus on fairness or equity does not seem to be a recent innovation. Indeed, Jones and Mann (2001) point out that concern for fairness and justice dates back to the origins of public-utility regulation:

   The concept of ‘fairness’ is deeply rooted in the 120-year history of public-utility regulation in the United States. Indeed, the word ‘fair’ appears in numerous regulatory concepts and propositions: for example, a fair rate of return, fair value, fair and reasonable rates, and full and fair evidentiary hearings. … While the idea of fairness is elusive and perceptions differ, it clearly is a potent force in regulation, as indicated by the vehemence with which participants complain when they feel they have been treated unfairly.  

Even in Australia it is straightforward to find explicit legislative requirements for the regulator to pursue fairness. For example, in the National Electricity Rules, the terms and conditions for access to negotiated distribution services and negotiated transmission services must be ‘fair and reasonable’ as must the price for connection to an existing network. An ‘efficiency benefit sharing scheme’ must provide for a ‘fair sharing’ of efficiency gains between a network provider and users.

Perhaps, however, these references to fairness merely represent a historical hangover from an earlier and less economically sophisticated time. Perhaps fairness concerns have diminished in recent years, especially following the recent reforms of the public-utility sector, with their focus on promoting competition? This doesn’t seem to be the case. In a survey of US state and federal regulatory commissioners, Jones and Mann (2001) found no indication that fairness is ranked lower than other goals of regulation than it was in the past. All but 4 per cent of the respondents ranked fairness as ‘either more important, or as

---

3 For example, Berg and Tschirhart (1995: 324): ‘...The fact is that most regulators do not view the world as economists view it. In reality, regulatory commission have objectives, motivations, and responsibilities far more complex than “setting price equal to marginal cost subject to a profit constraint”’. ... Neoclassical economics has focused almost exclusively on efficiency, which is but one of many objectives to most regulators. The justification for the narrow focus is that the other concerns, legitimate as they are, are too difficult to balance against efficiency given our regulatory control mechanisms. Whether or not this is true, it does not seem to dissuade regulators from a strong focus on equity or fairness.’

4 Jones and Mann 2001: 1. Bonbright, in his classic text of regulation, identifies fairness as ‘one of the three primary characteristics of a sound rate structure’ (Bonbright et al. 1988: 385).

5 Clause 6.7.1(9) and clause 6A.9.1(9). National Electricity Rules.

6 Clause 5.3.6, National Electricity Rules.

7 Muir (2001: 3) notes that: ‘Examination of the objectives of a small and fairly random selection of Australian and foreign utility regulators reveals that fairness figures explicitly in almost every case.’
important, as any other goal of public utility regulation’ (Jones and Mann 2001: 7). Sixty per cent reported that in a conflict between the goals of fairness and efficiency, fairness would be considered to be more important.  

**Dimensions of fairness**

There are many dimensions of what might constitute fairness, many of which go beyond the scope of this paper. To begin with, it is common in the literature on fairness to make a fundamental distinction between fairness in process, which is known as ‘procedural fairness’, and fairness in outcomes, which is known as ‘end-state fairness’ (see Konow 2003).

Procedural fairness (also known as ‘natural justice’) relates to the processes by which decisions are made. The principles of natural justice are designed to ensure that the decision-making process is fair and reasonable. I am primarily interested in the perceived acceptability of various pricing outcomes, rather than the process by which those outcomes are reached, so I will put to one side issues of procedural fairness.

There is also a sense of fairness which is linked with the distribution of income. This is sometimes known as ‘distributive equity’. Muir (2001) observes that:

> At the retail level, considerations of a customer’s ability to pay appear to dominate [considerations of fairness]. Underlying this concern appear to be attempts to improve the affordability of the service to lower-income consumers, those who are costly to supply or those with particular ‘merit’ claims (such as educational institutions or community organisations).

The linkage between income distribution and fairness is confirmed in laboratory experiments which show that providing information on the income levels of participants alters perceptions as to whether or not a given scenario is fair.

Nevertheless, as we will see later, fairness is more than just a surrogate for questions of income distribution. Fairness concerns remain even when participants can be presumed to have a similar income. For the purposes of this paper, I will follow the conventional economic approach of simply putting these questions to one side. In addition, for the purposes of this paper I will not make a distinction between the notions of fairness, equity, and justice. Instead, I will follow common usage and treat these terms largely as synonyms.

---

8 Jones and Mann (2001: 10). Zajac (1995: 101) and Baumol (1986) also argue strongly that economists should pay more attention to notions of fairness.  


10 The conventional economic justification for this approach is that ‘if one disapproves of the distribution of income, one might better correct it by lump-sum taxes (for example, on rents, income or inheritance) and
Is fairness systematic enough to be studied?

Historically, several authors have suggested we should not be over-optimistic about our ability to pin down just what constitutes fair, as opposed to unfair, pricing decisions. Fairness, it is argued, is an imprecise or elusive concept (Bonbright 1988; Baumol 1988; Fehr, Kirchsteiger and Riddl 1993), subjective (Konow 2006; Young 1994), or merely a cover for self-interest (Zajac 1995; Gielissen et al. 2008; Young 1994).

Nevertheless, a small but growing body of literature, starting with the pioneering work of Kahneman, Knetsch and Thaler (‘KKT’) (1986) has shown that there are empirical regularities in what members of the public perceive as fair or unfair in economic transactions. These regularities are discussed further in section 4. First, let’s review how economists have responded to fairness concerns in the past.

The response of economists to fairness objectives in public-utility regulation

When confronted with the fact that some of their most well-founded policy proposals are rejected by regulators on the grounds of fairness, the economics profession has reacted in different ways. It is possible to identify four different possible responses:

- The ‘conflict’ response, which asserts that the objective of conventional economic efficiency is paramount over all other objectives, and that any deviation reflects ignorance, incompetence, or capture by special interests.
- The ‘apartheid’ response, which asserts that objectives such as fairness or equity have a legitimate ‘separate but equal’ status to conventional economic efficiency, and the economist in his professional capacity has nothing to say about these other ‘non-economic’ objectives.
- The ‘assimilation’ response, which seeks to use economic tools and techniques to formalise or systematise the concept of fairness.

money transfers than by departing from the requirements of economically efficient pricing’ (Kahn 1970: Vol I: 68). This line of argument has been heavily criticised as not being economically sound. Nevertheless, it remains convenient for the purposes of this paper to put this aspect of fairness to one side.

11 Young (1994) opens his book on equity with several arguments that equity ‘fails to exist’: ‘The first is that equity is merely a word that hypocritical people use to cloak self-interest. It has no intrinsic meaning and therefore fails to exist. The second argument is that, even if equity does exist in some notional sense, it is so hopelessly subjective that it cannot be analysed scientifically. Thus it fails to exist in an objective sense. The third argument is that, even granting that equity might not be entirely subjective, there is no sensible theory about it, and certainly none that is compatible with modern welfare economics. In short it fails to exist in an academic sense.’
• The ‘integration’ response, which seeks to explain fairness and equity considerations as being derived from economic efficiency notions — possibly by varying the economist’s standard approach to efficiency.

The ‘conflict’ response

To many economists, the goal of economic efficiency (usually defined as allocative efficiency) is self-evidently paramount over all other objectives. Under this view, placing any weight on other objectives at best results in leaving untapped social welfare on the table.\(^{12}\) To do so reveals ignorance, incompetence, or capture by one or more interest groups.

Under this view, any departure from efficient policies, ostensibly due to concerns with fairness or equity, is ill-informed, misguided, a cover for self-interest, or at best secondary to the pursuit of allocative or productive efficiency.\(^ {13}\) Zajac (1995), for example, reflecting on his time at Bell Labs designing optimal rate schedules for AT&T, recalls that: ‘Those opposing us, we felt, were either misguided, ill-informed, or just plain evil.’ (Zajac 1995: 4).

Under this view, the task of the economist is to use the powers he has at his disposal to persuade the relevant authorities as to the virtues of maximising economic welfare (as conventionally defined), to expose the conflict with other objectives (see, for example, Bonbright 1988: 192), and to subordinate these other objectives to the efficiency objective.\(^ {14}\)

The ‘apartheid’ response

A second common reaction of economists is to declare that the objective of maximising conventional economic welfare is separate and distinct from the objective of promoting fairness or equity, and each has its own validity.

According to this approach, the discipline of economics is concerned with the narrow objective of promoting economic welfare (defined as allocative or productive efficiency). There may be other equally legitimate objectives, such as fairness or equity, but the economist has nothing to say about these objectives,

\(^ {12}\) Kaplow and Shavell (2002) go further to argue that placing any weight on a non-economic-welfare objective will result, in some circumstances, in leaving all members of society worse off.

\(^ {13}\) See Baumol (1986) and Bonbright et al. (1988: 182). The most active contemporary proponents of this view are Louis Kaplow and Steven Shavell of the Harvard Law School. See Kaplow and Shavell (2002).

\(^ {14}\) As an aside, there is of course, an equal-and-opposite view (not usually held by economists) that fairness and equity considerations must be paramount in the mind of the regulator, with allocative efficiency subordinate where it is considered at all. For example, Jones and Mann (2001) report that many regulators view fairness and equity considerations as more important than economic efficiency as conventionally defined.
at least not in his/her professional capacity as an economist. If regulators, serving their political masters, want to take into account other non-economic considerations, so be it.

For example, Bonbright asks whether or not the economist should ‘take part, in his or her professional capacity, in controversies about rival standards of fairness’. According to Bonbright, the usual answer is ‘no’, ‘on the ground that the question, being one of ethics, goes beyond their professional competence’ (Bonbright 1988: 192). Baumol is even more explicit:

The natural reaction of members of our discipline is to question the economists’ qualifications to make any pronouncements on fairness … Credulity is really strained by the prospect that an economist witness will tell a Congressional committee or a regulatory commission that he or she is qualified professionally to make some pronouncement on the fairness of some proposal that such a body may be considering. Why should economists ever make the attempt, and on what legitimate basis can they hope to do so? (Baumol 1987: 1)

Under this view, the economic and ‘non-economic’ objectives of a regulator are ‘separate but equal’. This response might be called the ‘apartheid’ approach.

The ‘assimilation’ response

A few economists have recognised the importance of fairness and equity considerations and have responded by attempting to formalise or axiomatise these concepts, so as to give the economist some ground on which to stand when providing advice on fairness issues or when trading-off fairness with conventional objectives.

There are two strands of this research worth highlighting. The first strand seeks to formalise the notion of fairness as equivalent to the ‘absence of envy’. As Konow (2003) explains, this concept was first formally stated by Duncan Foley (1967) and was further developed by Hal Varian (1974), Elisha Pazner and David Schmeidler (1978), William J. Baumol (1986) and others.

Under this approach, an allocation is said to be ‘envy free’ if no participant prefers the bundle of goods and services of another participant. This approach has the strong advantage that it uses many of the micro-economists’ existing tools and techniques, such as the concept of the Edgeworth box and the concept of the ‘core’ of a co-operative game.

Unfortunately, however, this approach has met with limited success in explaining the everyday concept of fairness. As Konow notes:
Absence of envy is an appealing construct and seems like a reasonable goal. The question asked ..., however, is whether it describes allocations people call fair, or whether it is distinct. Robin Broadway and Neil Bruce (1984) are sceptical about equating the two: 'I might envy a friend’s lucky find in an antique store yet perceive no “unfairness” that he, not I, owns it.'

Konow (2003) provides an example of a scenario in which the participants do not envy each other’s final allocation, but 90 per cent of respondents found the outcome unfair.

A second important strand of research starts from the observation that all tariff structures which yield the regulated utility sufficient revenue to cover its costs are equivalent to some allocation of the common costs — so the question as to what constitutes a fair or just tariff structure is equivalent to the notion as to what constitutes a fair or just cost allocation. This research then seeks to formalise the notion of a fair or just cost allocation.

This approach typically starts by asserting some fundamental principles or axioms that a fair cost allocation must satisfy, and seeks to characterise the class of cost allocations that satisfy this principle. The most well-known example of this approach is the famous work on cross-subsidies by Faulhaber (1975). Faulhaber starts from the principle that the revenue from any service or group of services should not exceed the stand-alone cost of this service. One possible justification for this ‘stand-alone cost ceiling’ is that it is ‘unfair’ for any group of customers to be paying more than what they could pay if they could organise to provide the service for themselves. As Ralph (1992) observes: ‘This seems a compelling minimum criterion for identifying a cross-subsidising group — common sense suggests it is unjust that a group should have to pay more than its stand-alone costs.’

There are important generalisations of this approach. Willig (1979) introduces the notion of ‘anonymous equity’. A cost allocation satisfies anonymous equity if no individual customer or group of customers is charged more than the stand-alone cost for the services that he/she (or the group) consumes. Young (1985) describes several other axiomatic approaches to cost allocation, including the ‘separable costs remaining benefits method’, the Shapley value and

---

16 Holcombe (1997) similarly rejects equating fairness with absence of envy (although in this case on procedural grounds).
17 In co-operative games a key question is how the gains from co-operation should be shared. The Shapley value, proposed by Lloyd Shapley in 1953, is a formula for sharing the gains from co-operation which satisfies certain key desirable axioms. See, for example, Young (1994), or Sergiu Hart (1989).
the nucleolus.\footnote{The nucleolus is a particular way of sharing the gains from co-operation introduced by Schmeidler in 1969. The nucleolus has certain desirable properties. For example, it is always unique and lies within the core (the set of all allocations for which no group of players could be better off if they refused to co-operate). See, for example, Young (1994).} Young suggests several properties that we might like a cost allocation to satisfy (specifically, ‘additivity’, ‘monotonicity’ and ‘consistency’). Not all of these properties can be satisfied simultaneously, but he shows how various combinations of these properties, with various ‘strengths’ lead to different cost-allocation methods.

To be clear, this line of research is independent from, and unrelated to, conventional economic efficiency notions. Zajac, discussing the various tests for cross-subsidisation notes: ‘These tests were … basically motivated by fairness, and we shouldn’t necessarily expect them to be economically efficient, nor that an economically efficient pricing structure will pass all of these tests.’ (Zajac 1995: 211).

But how well does this line of research do as a theory of fairness? Is it the case that, say, a random sample of objective members of the public would find cost-allocation methodologies which satisfy these axioms to be fair, while other cost-allocation methodologies which do not satisfy these axioms are deemed to be unfair?\footnote{Young (1994) observes that a ‘difficulty with the axiomatic method is that it can easily become disengaged from the problem that it was intended to solve. The invention of axioms and conditions is a fascinating business. The danger is that the exercise can take on a life of its own.’}

As we will see in the next section, empirical studies of fairness seem to show that the concept of fairness is not so much related to a particular tariff structure or cost allocation as it is to changes in that tariff structure or cost allocation. It seems that virtually any cost allocation could be considered to be fair if consumers have had a long enough period of adjustment. In other words, attempts to define what constitutes a fair cost allocation may be simply missing the point. As we will see when we survey the empirical literature on fairness, an allocation which satisfies relatively few axioms might be considered fair if customers have had time to adjust. Conversely, a move towards an allocation which satisfies several axioms (such as, say, the Shapley value) might, in fact, be deemed to be unfair — precisely because it raises the price on some group of customers.\footnote{As suggested later in this paper, the problem with this line of research is possibly that it focuses on too narrow a bargaining game between the firm and its customers. This paper suggests later that fairness is not simply a matter of sharing of the gains from co-operation at a given point in time; rather, fairness is about achieving efficient outcomes in a game in which one or both parties can make investments to increase the value of co-operation.} It seems that these approaches fail to capture key elements of public perceptions of fairness.
The ‘integration’ response

The fourth possible response to the persistent reference to fairness in regulation is what might be called the ‘integration’ approach. This approach seeks to explain the patterns of behaviour of regulators — including their description of some rate designs as fair or unfair — using economic models and approaches. This approach inevitably moves beyond the traditional focus on allocative efficiency and asks whether or not there are other economic models or approaches which can better explain the behaviour of regulators.

This paper falls into this fourth category. This paper seeks to identify patterns of regulatory outcomes that might be considered fair or unfair and then seeks to argue that these patterns can, in large part, be explained using economic models.

The next section attempts to identify patterns in attitudes towards fairness. The last section then sets out a model which seeks to explain these patterns in a manner that is consistent with economic efficiency.

What constitutes ‘fairness’ in public utility pricing?

The first step in this process is to identify key regularities in public notions of fairness. What, in fact, are the key characteristics of public attitudes towards fairness? Drawing on the few empirical studies on attitudes towards fairness in pricing more generally, I suggest that public attitudes to fairness in public-utility pricing can be summarised in the following five stylised facts:

Fact 1: It is considered unfair to raise prices above expected or anticipated levels

There appear to be solid grounds for the view that a key element of fairness relates to the stability or predictability of the regulated prices. It is perceived to be unfair to increase regulated rates above a promised, pre-announced, predicted, or anticipated level, except where the rate increase is necessary to offset an increase in long-term costs.

One of the key conclusions of the research by KKT is that the fairness of an economic transaction is measured in part by the change in the terms of the transaction relative to a ‘reference transaction’:

The main findings of this research can be summarized by a principle of dual entitlement, which governs community standards of fairness:
Transactors have an entitlement to the terms of the reference transaction and firms are entitled to their reference profit. A firm is not allowed to increase its profits by arbitrarily violating the entitlement of its transactors to the reference price, rent, or wage. When the reference profit of a firm is threatened, however, it may set new terms that protect its profit at transactors’ expense. (KKT 1986: 729–30)

KKT emphasise that it is the change in the terms of the transaction which is important, not any particular desirability of the original or reference transaction:

It should perhaps be emphasized that the reference transaction provides a basis for fairness judgments because it is normal, not necessarily because it is just. Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually … Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction. (KKT 1986: 730–31)

Interestingly, there is ample evidence from within the field of public-utility regulation that a change in regulated prices is often seen as unfair. Baumol (1988: 4) observes that: ‘It is important to recognize that “fairness” in a pricing arrangement depends heavily on consistency with the practices of the past to which people have become habituated.’

The same point is echoed by Kahn (1970, Vol I: 115): ‘As Ben Lewis has put it, “any scheme of compensation is fair provided only that it was reasonably anticipated at the time of the investment”.’

Along these lines, both Bonbright and Kahn go to great lengths to emphasise the importance of predictability and stability in regulated tariffs (Bonbright 1988: 387 and 187). A mail survey of a representative sample of Swiss households by Bruno Frey and Beat Gygi (1988, cited in Konow 2003: 1220) found that: ‘raising price in response to a demand increase is perceived as less unfair if the demand shift occurs at predictable intervals, alternative supplies exist, buyers are previously informed of and able to prepare themselves against the price increase, and sellers do not profit from the price increase’.

This is consistent with research by KKT which reveals that the permanence of the ‘reference transaction’ is a key determinant of the assessment of fairness. Specifically, it is not considered unfair to raise prices to eliminate a temporary discount or to cut wages by cancelling a worker’s bonus. For example, 61 per cent of respondents considered the following scenario unfair: ‘A small company employs several people. The workers’ incomes have been about average for the

21 See also Zajac 1985: 139, 141.
community. In recent months, business for the company has not increased as it had before. The owners reduce the workers’ wages by 10 per cent for the next year.’

On the other hand, 80 per cent of respondents considered the following scenario fair:

A small company employs several people. The workers have been receiving a 10 per cent annual bonus each year and their total incomes have been about average for the community. In recent months, business for the company has not increased as it had before. The owners eliminate the workers’ bonus for the year. (KKT 1986: 732)

As an aside, the perceived unfairness of price increases can (of course) lead to situations where prices do not rise to a level that is able to clear the market. KKT argue that this can explain the temporary incidence of rationing that arises from time to time. Several subsequent studies have confirmed that rationing is often perceived as more fair than raising prices to clear the market (see Raux and Souche 2006, and Dickson and Kalapurakal 1994: 431).

Fact 2: It is not unfair to raise prices to cover an unexpected increase in long-run costs

Although price stability is an important component of fairness, there are circumstances where price increases will be tolerated and considered ‘fair’ — in particular where the price increase is required to cover a change in the long-term costs of the supplier. KKT illustrate this with the following two questions. Seventy-nine per cent of respondents considered the following scenario was fair:

Suppose that, due to a transportation mix-up, there is a local shortage of lettuce and the wholesale price has increased. A local grocer has bought the usual quantity of lettuce at a price that is 30 cents per head higher than normal. The grocer raises the price of lettuce to customers by 30 cents per head.

Similarly, 75 per cent of respondents considered the following scenario was fair:

A landlord owns and rents out a single small house to a tenant who is living on a fixed income. A higher rent would mean the tenant would

---

22 ‘Conventional economic analyses assume as a matter of course that excess demand for a good creates an opportunity for suppliers to raise prices, and that such increases will indeed occur. The profit-seeking adjustments that clear the market are in this view as natural as water finding its level — and as ethically neutral. The lay public does not share this indifference. Community standards of fairness effectively require the firm to absorb an opportunity cost in the presence of excess demand, by charging less than the clearing price or paying more than the clearing wage.’ (KKT 1986: 735)
have to move. Other small rental houses are available. The landlord’s costs have increased substantially over the past year and the landlord raises the rent to cover the cost increases when the tenant’s lease is due for renewal. (KKT 1986: 733)

KKT summarise this result as follows: ‘[I]t is acceptable for firms to protect themselves from losses even when their transactors suffer substantial inconvenience as a result. The rules of fairness that yield such judgments do not correspond to norms of charity and do not reflect distributional concerns’ (KKT 1986: 733).

Interestingly, the previous two stylised facts could possibly be interpreted as suggesting that pricing according to the costs of providing the regulated service is considered fair, whereas pricing according to demand conditions is considered unfair. This, in turn, suggests that the common regulatory requirement that prices be ‘cost-based’ reflects an underlying concern for fairness.

**Fact 3: It is not unfair for the service provider to retain (or ‘share’) some of the benefits of a cost reduction**

Importantly, fairness does not require that any cost reductions are passed on, dollar for dollar, to users. Fairness allows the monopolist to retain a share of the benefits of any cost reductions it achieves. KKT (1986: 734) note that 79 per cent of respondents considered the following scenario to be fair: ‘A small factory produces tables and sells all that it can make at $200 each. Because of changes in the price of materials, the cost of making each table has recently decreased by $40. The factory reduces its price for the tables by $20.’

Allowing the regulated firm to keep some of the benefits of cost reductions could, of course, be necessary to induce investment in cost-reducing effort, in line with stylised fact 5 below.

In contrast, the KKT survey suggested that an attempt to retain all of the benefits of a windfall gain is considered unfair. Seventy-nine per cent of respondents thought the following scenario was unfair: ‘A grocery store has several months supply of peanut butter in stock…on the shelves and in the storeroom. The owner hears that the wholesale price of peanut butter has increased and immediately raises the price on the current stock.’

---

23 Dickson and Kalapurakal 1994: 431–2: ‘Overall these studies suggest that there will be a higher perceived fairness for cost-based rules in price determination, especially those that treat cost increases and decreases symmetrically. They imply that the traditional “fair” price is one based on cost, rather than market conditions.’

24 KKT (1986: 734) observe that this is consistent with a ‘dual entitlement view’ which ‘suggests that the firm is only prohibited from increasing its profits by causing a loss to its transactors. Increasing profits by retaining cost reductions does not violate the transactors’ entitlement and may therefore be acceptable.’
Fact 4: Even where it is unfair to raise the out-of-pocket costs of a service, it is not considered unfair to raise the opportunity cost of the same service.

In some cases a regulated firm may be able to, in effect, raise the ‘price’ of a service by lowering the price of the alternative. Importantly, it turns out that even where raising the price of particular consumption choice is considered unfair, it may be considered fair to raise the opportunity cost of taking the same decision.

For example, in the research of Raux and Souche (2003), peak-period pricing of TGV rail services or car-parking services was, as noted earlier, ‘overwhelmingly perceived to be unfair’. However, Raux and Souche also explored the possibility of rationing access to scarce capacity by ‘buying off’ those who were prepared to travel at another time (in the case of TGV services) or move to another car park (in the case of car-parking services). While only 10 per cent thought that rationing access to scarce TGV seats by raising the price was fair (whether the scarcity was one-off or recurring), 95 per cent thought that it was perfectly fair to ration access to scarce TGV seats by rewarding passengers holding reservations with a cash payment in exchange for switching to another day. This mechanism is, of course, similar to the practice of offering cash payments to airline customers to induce switching to another flight in the event of over-booking.

Similarly, while only 7–10 per cent thought it was fair to ration access to scarce car-parking spaces by raising the price, this increased to 89–93 per cent when the scarcity of car-parking spaces was rationed by ‘buying off’ customers willing to switch to another car park.

Fact 5: It is considered particularly unfair to raise the price for a service to exploit the good-faith effort or investment of the buyer in that service.

Importantly, the survey results suggested that respondents were particularly sensitive to the notion that the price might depend on certain characteristics specific to the buyer. Strikingly, more than 90 per cent of respondents thought the following scenario was unfair:

A landlord rents out a small house. When the lease is due for renewal, the landlord learns that the tenant has taken a job very close to the house and is therefore unlikely to move. The landlord raises the rent $40 per month more than he was planning to do.

Interestingly, the number of respondents who rated this outcome as unfair was 91 per cent — a higher ‘unfair’ rating than any of the other questions asked.
in the survey. Respondents clearly found this particular change in the terms of the transaction — which resulted from the decisions of the tenant — was particularly unfair. KKT observe: ‘The near unanimity of responses to this and similar questions indicates that an action that deliberately exploits the special dependence of a particular individual is exceptionally offensive’ (KKT 1986: 735).

In summary, I suggest that the key characteristics of fairness in pricing can be summarised in the following five stylised facts:

1. It is considered unfair to raise prices above reasonably expected or anticipated levels.
2. It is not unfair to raise prices to cover an unexpected increase in long-run costs.
3. It is not unfair for the service provider to retain (or ‘share’) some of the benefits of a cost reduction.
4. Even where it is unfair to raise the out-of-pocket costs of a service, it is not considered unfair to raise the opportunity cost of the same service.
5. It is considered particularly unfair to raise the price for a service to exploit the good-faith effort or investment of the buyer in that service.

Is there an economic-efficiency rationale underlying these notions of fairness? This is the question examined in the next section.

**Fairness and the promotion of sunk complementary investment**

The final step in the argument of this paper is to show that we can explain the empirical regularities set out in the previous section using a simple economic model. In particular, I suggest that, to a large extent, these patterns of fairness can be understood as an attempt to preserve incentives for valuable sunk investments on the part of both the monopolist and its customers.

**Sunk complementary investment and public utility regulation**

Although economic textbooks have, of course, long recognised that public-utility firms are often required to make a substantial long-lived investment, the possibility of sunk investment on the buyer side of the market has been largely neglected.
Conventional economic theory has treated buyers as largely passive and inert and has, for the most part, assumed that everything we need to know about a buyer is reflected in the demand curve. However, as emphasised in Biggar (2009), buyers are seldom completely inert and passive. Rather, buyers must often make their own sunk complementary investments to extract the maximum value from the monopoly service. The literature on transactions costs groups the different kinds of sunk investments into the following categories:

- The decision where to locate, when that decision will have an impact on the demand for monopoly services (for example, close to a rail spur, or close to a mine mouth, and so on). These are known in the transactions-costs literature as ‘site-specific investments’ (see Crocker and Masten 1996: 8).
- The decision to invest in discovering or developing new skills or knowledge when the value of that knowledge relies on a continuing supply of the monopolist’s product or service. These are known as ‘human capital-specific investments’.
- The decision to invest in sunk customer-premises equipment or other assets which are specialised to the monopolist’s product or service (such as electrical wiring on the customer premises). These are known as ‘physical asset-specific investments’.

To this list we might also add a fourth category: the decision to invest in the development and marketing of new products or services which rely on a continuing supply of the monopolist’s service. We might call these ‘product-specific investments’.

Where the benefits of these sunk complementary investments exceed the cost, it will be socially desirable for the investment to be carried out. However, buyers fear that, once they make these investments, the monopoly service provider will increase its prices, expropriating some or all of the value of that investment. This threat of hold-up has a chilling effect on buyer-side investment. The failure of the buyer to make that investment reduces the economic value of the monopoly service or, in some cases, prevents it being provided at all; in either case, reducing overall economic efficiency.

For example, in the absence of regulation or a long-term contract, a firm may be unwilling to construct an aluminium smelter supplied by a monopoly electricity-transmission grid for fear that the grid will increase its charges in the future. Similarly, a freight forwarder may be reluctant to purchase specialised railcars for fear that the price or quality of access to the rail track network will decline in the future. The failure to make these complementary investments represents a loss of overall economic efficiency.²⁵

²⁵ In some instances, of course, private arrangements can mitigate this hold-up problem. The clearest example of such an arrangement is a long-term contract between the monopoly service provider and its
Sunk complementary investment and fairness

Can this focus on the sunk investment of buyers explain the attitudes towards fairness in public-utility regulation identified above? We can make the following points:

First, this approach can explain the apparent ethical significance of the 'status quo' and the 'unfairness' of changes from the status quo. Buyers form expectations about the future path of tariffs and make their investment decisions accordingly. These expectations are typically formed by the current level and structure of tariffs, unless an element of the current level and structure is explicitly identified as temporary. A subsequent adverse movement in price or quality (relative to the expected level) is a potential threat to the value of that investment and therefore a potential threat to carrying out the socially valuable investment in the first place. By this hypothesis, therefore, an unanticipated adverse movement in prices will be judged as unfair. This is, of course, consistent with the first stylised fact identified above.

This link between ‘unfairness’ and sunk investments is highlighted by Meyer and Tye in their discussion of the difficulties associated with rapid rate changes following partial deregulation of the rail industry in the US. They note that where some consumers have:

sunk costs or made commitments that tie them for at least some time to particular vendors … the transition process can involve some aggrieved consumers who perceive themselves as unfairly victimized by the transition. The difficulty is created by an overhang of sunk costs committed under the prior rules of regulation. (Meyer and Tye 1985: 50)

Bonbright also notes an economic argument, based on sunk investments, that can explain the traditional focus on ensuring rate stability:

The argument runs to the effect that the ratepayers were induced to locate their factories, or to abandon their isolated generating plants, or to convert their furnaces from coal to gas burners, in contemplation of the low promotional rates and on the faith that that this rate would remain in effect for the indefinite future. (Bonbright 1988: 187)

At the same time, however, the promotion of rate stability could threaten the incentive of the regulated firm to make its own sunk investment, if the pursuit of rate stability prevented rates rising in response to a permanent increase in customers. However, in the case of many public-utility services the transactions costs of negotiating such contracts with customers render this approach infeasible. Instead, the public-utility regulator takes on the role of creating and enforcing the long-term contract that the parties would have written had they been able to negotiate before either party made any sunk investment. See Goldberg (1976); Biggar (2009).
costs. By this hypothesis, therefore, we would predict that a rate increase would be perceived as fair if it was necessary to cover a permanent increase in costs, consistent with the second stylised fact noted above.

On the other hand, an unanticipated reduction in costs poses no threat to the sunk investment of either the regulated firm or its customers. We might predict that it would not be labelled unfair to not pass on the full impact of a cost reduction to consumers. (Indeed, if the reduction in costs is brought about, at least in part, by cost-reducing effort on the part of the regulated firm, we might predict that it would be considered unfair to not allow the firm to share in some of the benefits of that cost reduction.) Overall, these predictions are consistent with the third stylised fact identified earlier.

Importantly, the mechanism used for rationing scarce supply can have a significant impact on the incentives for investment by buyers of the monopolist’s service. The decision by a consumer whether or not to make a sunk complementary investment involves a trade-off between the cost of the investment and the future increase in consumers’ surplus which that investment brings about. In the event that demand exceeds supply ex post (either due to an increase in demand or a shortfall in supply) the regulated firm must adopt some mechanism for rationing the excess demand. Where this excess demand is rationed by raising the price, the ex-post consumers’ surplus is reduced for all consumers — both those who choose to continue to consume and those who choose not to consume. This has a chilling effect on ex-ante investment. On the other hand, where excess demand is rationed by ‘bribing’ some consumers not to consume, the consumers’ surplus of all consumers is preserved — each consumer either continues to consume at the present price or is ‘bought off’, presumably when the reward for not consuming exceeds the consumers’ surplus from continuing to consume.

In other words, ex-post rationing through price results in a lower ex-ante surplus to the consumer and exposes the consumer to the risk of ex-post congestion compared to ex-post rationing by bribing some consumers not to consume. Unless the monopolist can directly subsidise the investment by the consumer and can make a credible commitment to maintain a given level of congestion, ex-post rationing through price will have a chilling effect on ex-ante complementary investment. By this hypothesis we would expect that unanticipated peak pricing would be considered unfair, while paying some consumers to not consume to relieve unanticipated excess demand would be considered fair, consistent with the fourth stylised fact set out above.

Finally, it is clear that raising the regulated tariff ex post on just those customers who have made a sunk complementary investment will have the effect of deterring further investment and therefore will be considered unfair. This
might arise, for example, if carrying out the sunk investment shifts the customer into a different tariff class with a higher price. By this hypothesis, these tariff structures would be considered to be unfair or ‘unduly discriminatory’, consistent with the fifth stylised fact above. At the same time, this approach allows us to suggest which forms of price discrimination would be considered fair — where customers of the regulated firm have had sufficient notice to adjust their own complementary investments to the pricing structure and where the tariff structure does not discriminate between customers on the nature or extent of their sunk complementary investments.

Similarly, this approach predicts that peak-load pricing could, under some circumstances, be considered acceptable — particularly when customers have sufficient notice to adjust their own complementary investments, where the frequency of the peak periods is fixed in advance, and where the peak prices themselves are fixed in advance. There is some evidence to support this. Raux and Souche (2006) have shown that people are less averse to pricing in a recurring situation where there is time to adjust to the price fluctuation than in an exceptional situation where pricing serves to ration demand.

In summary, I suggest that the notion of fairness in public-utility regulation — and in public pricing more generally — can be explained as an attempt to protect and thereby promote sunk complementary investment by users and consumers. I propose that regulatory decisions consistent with the hypothetical efficient ex-ante long-term contract between the regulated firm and its customers will be considered fair, while regulatory decisions which are inconsistent with that hypothetical ex-ante long-term contract will be considered unfair.

Alternative explanations?

Is there an alternative approach which can better explain the pattern of attitudes to fairness set out above? For example, some economists might argue that claims of fairness are merely a veil for the pursuit of self-interest. Perhaps regulators, in pursuing fairness over traditional economic efficiency, are merely bowing to the demands of powerful interest groups? As noted earlier, some research has indeed found that fairness judgments are biased by self-interest.

But fairness seems to be more than a mere reflection of self-interest. The research of KKT (amongst others) shows that disinterested, objective observers hold consistent notions of fairness even when they have nothing at stake. If fairness is nothing more than a reflection of the self-interest of consumers, why would it be considered fair for service providers to retain some of the benefits of a cost reduction (Fact 3)? If fairness is a reflection of the self-interest of the regulated firm, why would it routinely be held to be unfair to raise prices above expected
or anticipated levels (Fact 1)? The self-interest hypothesis seems to have trouble explaining these aspects of attitudes towards fairness. Fairness, as I have sought to argue, seems to be something more than a mere reflection of self-interest.

What about the axiomatic approaches to fairness discussed earlier? As noted earlier, certain cost-allocation methodologies, such as the Shapley value, can be justified as satisfying certain desirable axioms. But, as noted earlier, this line of research seems to do a poor job of explaining public attitudes to fairness. Specifically, as already noted, fairness seems more closely related to changes in allocations than to the allocations themselves.

We are left with the hypothesis put forward in this paper — that fairness concerns can be explained as an attempt to protect and promote the sunk complementary investments of the regulated firm and its consumers. Does this hypothesis have any testable implications? One possible implication is that concerns of fairness or unfairness would be expected to arise more frequently in those sectors in which customers have made a material sunk investment in a relationship with the supplier. As noted above, sunk investments on the part of customers are reasonably common. However, those sunk investments are not specific to a particular supplier precisely in those markets where the customers can switch easily between suppliers — that is, in those markets which we would normally label competitive. Therefore, an implication of this hypothesis is that we would expect to see concerns of fairness or unfairness arise much more frequently in those sectors in which customers have no choice of supplier (that is, in public-utility industries) than in those sectors which are reasonably competitive (and in which customers make sunk investments specific to an individual supplier). This implication seems broadly consistent with casual observation.26

However, even where customers face some choice of supplier they may be required to make some degree of relationship-specific investments. Where those investments cannot be adequately protected through long-term contracts, we might expect to see fairness concerns, and some of the pricing outcomes discussed here, arising in normal commercial transactions. Blinder et al. (1998), in a study of price stickiness, find that around two-thirds of private, non-farm, unregulated, for-profit US firms responded ‘yes’ when questioned whether firms have an implicit understanding with their customers, who ‘expect the firms to not take advantage of the situation by raising prices when the market is tight’. Interestingly, consistent with the results reported above, Blinder et al. find that

26 Another observation we might make is that both firms and households must make sunk complementary investments as customers of a regulated monopoly firm. Therefore another implication of this thesis is that we should expect to find no material difference in fairness concerns when the primary customers of the monopolist are other firms compared to the case where the customers are households (except where the customer firms are sufficient large and are present before any investment by the monopolist has been sunk, so that the negotiation of long-term contracts is a viable alternative).
these implicit contracts tolerate price increases when costs increase but do not symmetrically insist on price reductions when costs decrease. ‘Furthermore, the notion that prices should be stabilised seems to apply a bit more to demand shocks than to cost shocks’, confirming the suggestion by Okun (1981) that ‘price increases that are based on cost increases are “fair” while those based on demand increases are often viewed as “unfair”.’ These findings are consistent with the discussion in this paper. The central claim of this paper — that fairness in public-utility pricing can be understood as an attempt to protect and promote sunk investments — may extend to commercial pricing practices more generally.

**Conclusion**

Economists have long viewed users and consumers as essentially passive — so that all the relevant behaviour of consumers can be summarised in the demand curve. In this framework, the most important goal of the regulator is to minimise the deadweight loss. In fact, users and consumers can, and often must, make substantial sunk investments to extract the most value from infrastructure assets. The lack of recognition of these investments in mainstream regulatory policy has contributed to a lack of understanding and conflict between the allocative-efficiency aims of economists and the fairness notions of regulators. I suggest that the desire to protect and thereby promote these sunk complementary investments can explain much of the notion of fairness as it applies to public-utility regulation and public pricing more generally. In my view, by taking into account the impact on these sunk investments in their policy advice, economists will go some distance to bridge the remaining gap between economic theory and regulatory reality.

**References**


Willig, R. D. 1979, ‘Customer Equity and Local Measured Service’ in Joseph A. Baude et al. (eds), Perspectives on Local Measured Service, Kansas City.


