Capital Markets that Benefit Investors

A Survey of the Evidence on Fragmentation, Internalisation and Market Transparency

Ruben Lee

Oxford Finance Group
30/9/2002
This report was financed by: the Association of Private Client Investment Managers and Stockbrokers/European Association of Securities Dealers; British Bankers Association; Danish Securities Dealers Association; Futures and Options Association; International Primary Market Association; International Securities Market Association; International Swaps and Derivatives Association; London Investment Banking Association; Swedish Securities Dealers Association; The Bond Market Association; with additional support from the Corporation of London.
Acknowledgements and Disclaimer

The author would like to thank Bruno Biais, Lawrence Harris, Maureen O’Hara, Ananth Madhavan, and Hans Stoll, for permission to reproduce some text from their respective works, and Robert Battalio, Ryan Davies, Alfonso Dufour, Ananth Madhavan and Charles Sutcliffe for their help in preparing the report. The views expressed here are those of Ruben Lee alone and do not necessarily represent the views of any of the sponsors of the report. While every effort has been made to ensure the accuracy of the report, no factual material or statements presented here are guaranteed correct.

Author’s contact email address: rubenlee.ofg@btinternet.com
Abstract

This report examines the extent to which existing published evidence answers key questions about whether competition/fragmentation and internalisation harm market performance or best execution, and about what effects various levels and types of transparency have on market performance and best execution. The evidence is complex and supports different and sometimes contradictory conclusions. On balance, it shows that competition between exchanges and different trading venues enhances market performance and benefits investors. While the early evidence identifies the costs of internalisation, more recent empirical and theoretical studies highlight the benefits that internalisation brings to investors. Although greater transparency improves the extent to which, and the speed with which, information is incorporated into share prices, there is growing evidence that too much transparency may harm market performance.
Executive Summary

Introduction

The aim of this report is to provide a brief and understandable summary of the economic evidence on three topics that are crucial to the current debate about which securities market trading structures will best enable Europe to achieve its economic goals. These three issues are fragmentation, internalisation, and transparency. The report summarises the current state of research into how these phenomena affect market performance and best execution.

Three introductory comments on market performance and best execution are important. First, the quality of a market with many trading venues is dependent on how the various criteria determining market performance are assessed over the whole market, and not just across any single trading venue. Second, economics, by itself, is not able to assess the fairness of a market, although it can provide an assessment of the costs of choosing one set of values over another. Third, many aspects of a trade, in addition to its price, may be relevant for best execution. As different traders have different trading preferences, no single definition of best execution is possible - it depends on what is important to the relevant investor.

A. Fragmentation

Apart from a few notable exceptions, there is now a wide consensus that it is beneficial to have competition between exchanges, and other types of market participants, for the trading in a particular security. Seven main and closely related themes have been raised in the debate over whether all orders in a particular security should be consolidated on a single trading system, or whether orders should be allowed to fragment to competing trading systems. Each theme is discussed in turn.

1) Competition between Exchanges and Other Types of Trading Systems: Such competition keeps the services which exchanges and trading systems offer at a low price, including transaction, clearing and listing services, and provides an incentive for them to innovate and enhance their product lines, for example by automating various aspects of their systems and trading different assets.

Investors go where they believe their orders have the greatest chance of being executed. Order flow thus tends to attract order flow, which means that trading in any particular security is likely to be concentrated on a single dealing mechanism, or at most a small number of competing systems. If, however, trading is fragmented between different trading venues for whatever reason, it has been argued theoretically that this dispersal of orders lowers the probability of execution at each location and may therefore reduce liquidity, and also increase the volatility of transaction prices. Other theoretical analysis of the competition between exchanges identifies the possibility that
different trading systems may co-exist, if traders have diverse needs and desires, and if each trading system satisfies the trading preferences of a particular group of investors. Some theoreticians show that competition between trading systems is not always beneficial, and that it can sometimes decrease, as well as increase, liquidity.

In the European context, a range of empirical analyses have been made of the competition between the national exchanges following the Big Bang in London. Notably, this period was characterised by a bout of intense innovation at the various exchanges concerning their market structure, their technology, and the services they offered. Most of the measures of market performance - for example, in creating a diversity of trading venues, in linking the markets, and in being able to deal in greater size at the available spreads - show that such competition was beneficial.

A large amount of empirical research on the effects of competition between trading systems has also been undertaken in the US. Much of this work concludes that increased competition in securities markets leads to tighter bid-ask spreads, and with some exceptions, has also had a range of other beneficial effects.

2) **Different Preferences of Different Traders**: The existence of different trading systems with distinct market architectures brings choice to investors. Different benefits are available from different types of trading systems. The transaction costs of demanding immediacy of execution, for example, can be significantly higher than the costs incurred in trading patiently. Large cost savings can also be achieved by using automated trading systems instead of traditional intermediated markets.

The existence of multiple trading venues, and the diversity amongst the preferences of investors, has put pressure on what best execution really should mean. Contrary to the historically accepted approach, some recent research suggests both that the concept of a best price is not well defined in the context of multiple trading systems, and that attempts to mandate best execution as a consumer-protection device can limit competition between trading systems, and thereby not give the result that the consumer wants.

3) **Inter-Dealer Competition**: Such competition keeps the services which dealers offer at a low price, provides an incentive for them to enhance their products, and can also reduce the market power of price-setting agents and thus result in better execution conditions.

4) **Price Priority**: With price priority, the highest bids and lowest offers get executed before other orders. A lack of price priority may lead to a reduction in the speed and accuracy with which information about a company is incorporated into the price of its shares, and in the quality of price formation of a market. Although it may be more difficult to deliver price priority in a market with
competing trading systems than in one centralised trading system, market participants themselves have an incentive to ensure that they receive the best price, and thus that price priority obtains.

In order to understand much of the evidence about fragmentation, it is important to appreciate the difference between two types of orders: so-called ‘informed orders’ and ‘uninformed’ orders. ‘Informed’ orders are said to come from market participants who have some knowledge about the future value of a security they are trading, whereas ‘uninformed orders’ are from market participants who have no special information about the future value of the security. The execution of uninformed orders presents a lower risk to an intermediary than the execution of informed orders. This is because when an intermediary trades with an informed order, the likelihood is that the price of the security will move in favour of the informed trader, and against the interest of the intermediary. The dealer is thus said to face an ‘adverse selection’ problem when trading with informed traders.

Much early empirical work concludes that the NYSE offers better prices than those available on other trading venues. This evidence has been used to suggest that trades executed off the NYSE may therefore not obtain best execution, in terms of receiving the best available execution prices. It has also been widely argued that payment for order flow diverts low risk ‘uninformed’ orders away from the primary trading venue, and leaves it only the more risky ‘informed orders’ for execution. On this analysis, “cream-skimming”, as it has been called, may undermine the process of price discovery because the diverted orders do not contribute to the price discovery on the primary market. Furthermore, if quote matching is employed, diverted orders may be executed at worse prices than would have occurred if those orders had been entered on the primary market. There is, however, growing evidence suggesting that market quality is not any worse when trading takes place off the primary market, and furthermore that the quality of such executions, when factors other than price are taken account of, may indeed be better.

A key difficulty in assessing the costs and benefits of fragmentation is that competition between orders may work best in a single consolidated market on which all orders can interact against each other, whereas competition between exchanges, trading systems, and OTC trading, and all the benefits such competition yields, implies fragmented markets. The two forms of competition may not be compatible with each other.

5) **Secondary Priorities**: Time priority gives market participants an incentive to submit limit orders early to a trading system, because by doing so they are more likely to have their orders executed. Time priority is difficult to enforce in a world of competing trading systems. It has been suggested, theoretically, that a lack of time priority may lead spreads to widen, and may also reduce the chances of limit orders being executed. Advances in automation may, however, decrease the importance of having an enforceable secondary time-priority rule. This is because a similar outcome
can be achieved by monitoring the market closely, and by submitting a market order should the price of the relevant stock move to where the limit would previously have been placed. This strategy also minimises the costs to investors who submit limit orders of providing a free option to other investors to trade against these limit orders when prices move against them.

6) **Consolidation of Information:** The consolidation of information means that quote and trade information about the trading in a particular security is readily available from a single source. It is sometimes argued that the consolidation of such information is easier in a consolidated market, although private solutions to consolidating such information across markets have also been developed in many contexts.

Transparency is widely considered to play a key role in consolidating markets, by ensuring that prices are equalised over all trading venues. It allows arbitrageurs to trade whenever prices on one trading system are inconsistent with prices on another trading system. The view that full transparency is required in order to ensure price equalisation across trading systems is not, however, universally accepted.

7) **Public Good Nature of Various Exchange Services:** Such services include price stabilisation, market surveillance, self-regulation, and quality certification. The existence of multiple centres for trading an asset may reduce the provision of the public goods that exchanges provide, given the difficulty of funding them. It may also be easier to share regulatory costs on a fair basis in a consolidated market. However, various regulatory solutions have been proposed and implemented to address the problems of ensuring that trading on all available venues is adequately regulated, and that such regulation is funded appropriately.

**B. Internalisation**

Given that ‘internalisation’, the execution of trades in-house rather than on an exchange’s order book, is a form of fragmentation, many of the themes discussed in the context of fragmentation are repeated here. The early analysis of internalisation stressed two key potential costs associated with it. First, internalisation was shown, theoretically, to have the potential to harm the price discovery function of a market, because it stops all orders in a particular security from competing directly against each other on a single order book. Such fragmentation, it was said, might also lead to larger bid-ask spreads and greater price volatility.

A second key problem identified theoretically with internalisation is that it may divert ‘uninformed’, low risk, trades away from a primary exchange, leaving only the more ‘informed’, and higher risk, orders on this primary exchange. This ‘cream skimming’ may lead dealers to set wider spreads on the primary exchange in order to protect themselves from being picked off by investors with better information. If the diverted ‘uninformed’, and low risk, orders are executed at
prices taken from the primary market, they may also then be executed at worse prices than if they had been entered on the primary market.

Two key benefits with internalisation have subsequently been noted. First, it might allow alternative trading venues to compete with a primary or central market, and such competition could lead to enhanced pricing. Second, it would give market participants a greater diversity of choice of where to execute their orders, rather than having only the single option of sending them to the central market. This might benefit particular classes of investors. For example, the execution of retail traders’ orders via internalisation may let them obtain better prices and lower dealing costs than on a primary exchange, if their orders are low risk ‘uninformed’ orders. If dealers compete for retail orders, they will set smaller spreads for these low risk uninformed orders than would occur in a consolidated market, given that the dealers do not have to protect themselves against being picked off by dealing with investors with better information.

In the section on fragmentation, it was shown that much of the early empirical work on internalisation (almost exclusively about the US markets) concluded that the primary exchange (the NYSE) offers better prices than those available at other trading venues. A growing body of more recent evidence about internalisation and the practice of some trade execution venues in the US of paying brokers to execute their orders at the execution venue (known as ‘preferencing’), however, highlights the benefits available from internalisation, and questions both whether better prices are in fact available from the NYSE, and also more generally whether better execution is available. Using various different measures, internalisation and ‘preferencing’ have been shown not to harm the execution of market orders or limit orders, and may indeed improve it. One example is that market orders traded via ‘preferencing’ on regional exchanges tend to trade more favourably relative to the NYSE, than market orders placed on ‘non-preferencing’ regional exchanges; another is that limit orders have a greater probability of executing on regional exchanges than on the NYSE. Internalisation also seems to have little short-run effect on posted or effective bid-ask spreads.

There is conflicting data about whether internalisation leads to cost competition or to the problem of ‘adverse selection’ noted above, when a dealer is likely to lose money, and therefore set higher spreads, if he deals mainly with informed traders. The most recent theoretical analysis of how internalisation functions takes account both of the prices that investors receive and of the commissions they pay for their executions. The conclusions about the effects of internalisation on best execution are more complicated than previous analysis suggests. In particular, it is shown that payment for order flow can lead both to an increase in execution quality, taking into account both the price and commissions paid, which is consistent with cost competition. At the same time, however, it can give rise to an increase in the proportion of higher risk informed trades on a primary exchange, which is consistent with cream-skimming.
There is empirical evidence that when best execution takes account of factors other than merely price, the internalisation of orders can lead to better executions than are available on the NYSE. For example, while execution on the NYSE appears better than internalisation using measures of trade-price quality, internalisation provides more timely executions and produces more liquidity enhancement than at the NYSE. If potential reductions in commissions are taken into account as well as execution prices, previous results showing that better execution has been available on the NYSE than elsewhere may need to be revised.

A range of theoretical reasons have been put forward for why institutional traders may like to use a particular form of internalisation known as the ‘upstairs’ market (a US term referring to the activities of members of an exchange, namely brokers and dealers, when they search for counterparties for big institutional orders off the floor of the exchange). Upstairs markets may help institutional investors locate trading counter-parties, and thereby execute large trades without fully revealing their orders to the ‘downstairs’ market. ‘Upstairs’ counter-parties may be able to filter out large ‘informed’ orders. Upstairs trading may facilitate the collection of information about the unexpressed supply of, and demand for, securities. Upstairs trading may also help risk-sharing amongst market intermediaries, thereby lowering transaction costs.

Empirical evidence from a range of different contexts shows the benefits of upstairs trading, which occurs not only in the US, but also in many other countries including in Europe. On the NYSE, upstairs trading appears to be used by investors who can signal credibly that their trades are ‘liquidity’ motivated, and not motivated because they have better information. The upstairs market may thus enable transactions that would otherwise not occur in the downstairs market. In Australia, off-exchange trading - including ECN, upstairs and after-hours trading - is shown to benefit those traders in a position to switch trading venues, by lowering their trading costs. In Canada, the upstairs market makes on the Toronto Stock Exchange provide a vehicle for screening out orders from investors who have better information, and for executing large liquidity motivated orders at a lower cost than the downstairs market.

C. Transparency

The key effects of transparency on securities markets, and on the participants in such markets, are complex and contradictory.

Almost all the evidence shows that greater transparency improves the speed and accuracy with which information about a company is incorporated into its share price. It may allow traders to select which trading system delivers the best quoted price, thus facilitating arbitrage between different systems, ensuring price priority, and enhancing the price discovery process. In turn, this is often believed to enhance best execution.
There is growing evidence, however, that greater transparency may also harm market performance in various ways. Not all investors are willing to expose their orders publicly, given the risks they run by doing so. Enhanced transparency can reduce their willingness to participate in the market, and has been shown to decrease liquidity in various contexts.

Traders with superior information are likely to prefer trading systems with less pre-trade transparency, so that they can keep confidential their trading intentions. However, uninformed traders, with no particular information advantage, prefer greater transparency. No single transparency regime will therefore be seen as optimal by everybody.

Transparency may make bid-ask spreads widen because market makers have less incentive to pay to capture the information that a trade with an informed trader will bring. Alternatively, spreads may decrease because information about transactions reaches all market participants, or because dealers are aware of each others’ positions and compete more strongly with each other.

Transparency may encourage stabilising speculation that helps absorb order flow imbalances between buy and sell orders, and reduce volatility. Alternatively it may exacerbate market participants’ strategic behaviour towards each other, with the possibility of increasing volatility.

The level of transparency in a market also affects how trading systems compete against each other. Several commentators suggest that trading systems should own the property rights in the information arising from their trading systems, and that they have the appropriate incentives in most circumstances to determine the appropriate levels of transparency for their trading systems.
Introduction

The aim of this report is to provide a brief and understandable summary of the economic evidence on three topics related to market structure that are crucial to the current debate about which securities market trading structures will best enable Europe to achieve its economic goals. The three issues are fragmentation, internalisation, and transparency. The report summarises current knowledge about how these phenomena affect market performance and best execution. It is composed of four main sections, and three appendices. In the first section, some background information on the concepts used in the report is provided. Section two contains a broad overview of the literature on the general topic of fragmentation. The effects of internalisation, one important kind of fragmentation, are summarised in section three. In section four, the implications of transparency for market performance and best execution are described. The three appendices contain, respectively, the notes to the text, some excerpts from other literature summaries on the topics at issue,\(^1\) and a list of the references employed in the report.

1. Conceptual Background

Given their importance, and also the confusion and controversy which sometimes surround their use, the following concepts are briefly explained in this section: market structure, market performance, liquidity, best execution, fragmentation, internalisation and transparency.

The **market structure** (or architecture) of a single trading system normally refers to the rules governing how the trading system delivers the three functions of data dissemination, order routing, and order execution. Data dissemination is the act of transmitting pre- and post-trade data, about quotes (or orders) and trades respectively, to market participants. Order routing is the act of delivering orders from their originators, such as investors and financial intermediaries, to the execution mechanism. Order execution is the process whereby orders can be transformed into trades. Order execution is sometimes called “matching” or “clearing”. Sometimes the term market structure is used to refer only to a trading system’s order execution algorithm. The actual, let alone potential, diversity of market structures across different trading systems is enormous.\(^2\) While a few of the new types of trading systems - including Alternative Trading Systems (ATSs), Electronic Communication Networks (ECNs), and crossing networks - have experimented with market structures not previously employed by traditional exchanges, many have simply replicated traditional auction market structures, but in an automated manner.

In a market with more than one trading system, the structure of the market as a whole normally refers to the combination of both the individual structures of each trading system in the market, and to the linkages, be they formal or informal, that connect the various trading systems in the market together.
**Market performance** is a measure of how well a market performs. When assessing the quality of a particular market, it is the market’s performance, and not its structure that is important. The following non-exclusive attributes of a market are normally considered central in evaluating its performance:

- *Price discovery* (or price formation) - the extent to which, and the speed with which, information is incorporated into prices. A market in which information is fully aggregated into prices is called “informationally efficient”, “price efficient”, or sometimes simply “efficient”.
- *Liquidity* (discussed at greater length below) - a measure of the extent to which market participants can rapidly execute large-volume transactions with a small impact on prices.
- *Volatility* - a measure of the extent to which the prices of the assets traded in a market vary, both individually and collectively.
- *Allocational efficiency* - the extent to which the market allocates resources to their most productive uses.
- *Efficiency* - the extent to which transaction costs in general are minimised.
- *Integrity* - the extent to which the market is safe, sound, and free from systemic risk.
- *Investor protection* - the extent to which market participants are protected from fraud and abusive activity.
- *Fairness* - the extent to which market participants are treated in a just and non-discriminatory manner.

Five aspects of these attributes are noteworthy: 1) Many of them are difficult both to define and measure. 2) The realisation of more than one of them may complement each other in some circumstances, but in others may conflict with each other. 3) The quality of a market with many trading venues will be dependent on how these attributes are assessed over the whole market, and not just across any single trading venue. 4) The economic analysis of the effects of market structure on market performance focuses for the most part on the first three of these attributes, namely price discovery, liquidity and volatility. 5) Economics, by itself, is not able to assess the fairness of a market. It can, however, provide an assessment of the costs of choosing one set of values over another.

**Liquidity** is a multi-dimensional concept. At the most general level, a liquid market is one in which participants can rapidly execute large-volume transactions with a small impact on prices. Four dimensions of liquidity are typically identified: width, depth, immediacy and resiliency. The width (or tightness) of a market is a measure of how far quotes or trade prices diverge from some measure of the “true”, or full-information value of a security. The width of a market is often assessed using different measures of either quoted bid-ask spreads, “effective” spreads - which are
typically calculated as the difference between transaction prices and contemporaneous mid-quote prices, or “realised” spreads - which are typically calculated as the difference between transaction prices and subsequent mid-quote prices. Depth is a measure either of how much prices need to change in order to execute larger sized transactions, or the amount of orders on an order book, or on the books of market-makers, at the minimum spread at a given time. Immediacy is a measure of how rapidly a trade can be executed. Resiliency normally refers to the speed with which prices revert to their “true” values after a trade. The assessment of the liquidity in a market is controversial, and many different measures are possible.³

Four key factors have been identified as giving rise to liquidity costs, and they are briefly noted here, because much of the literature examined assumes a knowledge of how these costs operate. The first source of liquidity costs is order-processing costs which are typically incurred by dealers in processing orders. These are normally assumed to be a fixed value per trade, independent of its size. A second source of liquidity costs is inventory or risk-sharing costs. When a dealer purchases a security from a market participant, he incurs a risk that the value of the security may change. He thus needs to be compensated to take on this risk. A third source is search costs. In order for a market participant to trade a security, he needs to find a counter-party on the other side of the market willing to take the trade. This search process can be expensive.

A final factor that may give rise to liquidity costs arises from the presence of asymmetric information. Such costs are thought to arise because of a difference in the motivations for trade between two types of investors: one group called “liquidity”, “noise” or “uninformed” traders, and the other group called “informed” traders.⁴ Liquidity traders are assumed to trade securities for reasons that are exogenous to the future value of the security, while informed traders, in contrast, are assumed to trade because they know something about the future value of the security. Whoever an informed trader trades with is likely to lose money, given that the informed trader is trading only because he knows something about the future value of the security. If an intermediary, for example, deals with an informed trader, the informed trader is likely to make money at the expense of the intermediary. The problem for the intermediary, however, is that typically he cannot distinguish between liquidity traders and informed traders, and the informed traders know this. As a result, the intermediary may need to charge a fee to protect himself against the adverse selection costs that he faces because informed traders choose to deal with him.

**Best execution** is a measure of how well investors’ trades are executed. The concept is particularly relevant when evaluating the performance of an agent in the market, such as a broker, who acts on behalf of customers in facilitating the execution of their trades. The price of a trade is typically the most important component of best execution. However, many other aspects of a trade may also be relevant, including execution speed, the opportunity for price improvement, the
probability an execution will be achieved, the maintenance of trader anonymity, opportunity costs, price impact, commissions, and the minimisation of other trading system charges.

The notion of best execution has become controversial for several reasons. Four are noted here. First, as different traders have different trading preferences, no single definition of best execution is possible - it depends on what is important to the relevant investor. Second, given that traders and investors frequently assess their execution against multiple criteria, some of which may be mutually exclusive, a method of ranking the various criteria is required. This is not easy to obtain. A third difficulty may arise, because it is not easy to measure many of the aspects of best execution. Finally, another reason why it is hard to measure best execution is that it requires measuring not only the attributes of the trade in question, but also all other relevant opportunities that are reasonably attainable in the market at the same time. The issue of what is reasonably attainable is not easy to determine.

**Fragmentation** is normally said to occur if the supply of goods or services in a market is divided in some manner, typically because there is no single or dominant supplier operating in the market. A competitive market requires by definition that there be more than one supplier in the market, and as such competitive markets are always fragmented. In the context of securities markets, fragmentation arises when all orders do not interact with each other via a single order execution mechanism. This may happen if investors send their orders to more than one trading venue, be it an exchange or other type of trading system, and it can also occur through the process of internalisation.

Confusion may arise about the meaning of the term fragmentation because it is sometimes used not only to describe the structure of a market, but also to imply that the performance of a market is impaired as a result of such a split or division. In this report, the term is taken to be applicable solely to market structure and not to market performance. A key question, then, is does competition/fragmentation harm either market performance or best execution.

**Internalisation** occurs when a broker acts as a dealer, or market maker, to execute his clients’ orders, without exposing them to an exchange or primary market. The term internalisation is also sometimes used more broadly to refer to trading on market systems which, although they do not involve exposure of orders directly on a regulated market, do provide an open and transparent market architecture. Internalisation is a type of preferencing. Preferencing occurs when a broker-dealer routes his order flow to a specific dealer, market maker or trading system, to take priority in execution over same-priced orders or quotes entered prior in time. Such order flow routing typically arises either because there is some form of affiliation between the broker and the dealer, or because the dealer pays the broker in some manner to receive the broker’s orders. One important type of internalisation is when an intermediary crosses large trades off the order book of a
continuous auction exchange. The possibility of such “upstairs” trading is sanctioned by many exchanges in their rule books, including the major European exchanges.

**Transparency** refers to the extent to which information about quotes and trades is publicly disseminated. “Pre-trade” transparency refers to whether information about the prices and quantities of quotes are publicly disseminated immediately after they have been submitted to a trading system or exchange. “Post-trade” transparency refers to whether information about the prices and quantities of trades are published immediately after they have been executed. The “reporting” of quote and trade information is used here to refer to the act of reporting it to the relevant regulatory authorities. Reported information may or may not be published subsequent to having been reported. In the US, the phrase “last-sale reporting” normally means post-trade transparency.

There are substantial differences between the types of data about quotes and trades that trading systems both can, and choose, to release. The choice by a trading system or exchange of what quote and trade information to release, namely what level of transparency to adopt, is a central element of the wider decision as to what market architecture to employ. Almost all exchanges and trading systems disseminate some quote and trade information in order to attract trading. Not to do so would make it hard to attract order-flow, given that market participants will not normally send their orders to a trading arena without some knowledge of the quotes and prices arising from it. The strategic non-disclosure of some information about quotes and trades is, however, also a central element of all markets’ architectures. For example, many exchanges allow the placing of “iceberg orders” which permit investors to reveal publicly that they desire to trade in only a small number of shares, while still maintaining an order for a larger amount of shares on the order book.

2. Fragmentation

This section contains an overview of the key themes that have been identified in the voluminous literature on the general topic of fragmentation. In addition, and given the widespread attention that is currently being paid to new types of trading systems (such as ATSSs, ECNs, and crossing networks), a more technical and detailed summary of the effects that competition by such trading systems is having on market performance and best execution is also presented.

2.1. Key Themes

Apart from a few notable exceptions, there is now a wide consensus that competition between exchanges, and other types of market participants, for the trading in a particular security, is beneficial. The rules governing how such competition should operate, however, remain controversial. Seven main and closely related themes have been raised in the debate over whether all orders in a particular security should be consolidated on a single trading system, or whether orders should be allowed to fragment to competing trading systems [following Ch. 8, Cohen, Maier,
Schwartz, & Whitcomb (1986), Ch. 13, Schwartz (1988), and Ch. 9, Schwartz (1991). They are as follows: 1) competition between exchanges and other types of trading systems, 2) different preferences of different traders, 3) competition between dealers, 4) price priority, 5) secondary priorities, 6) consolidation of information, and 7) the public good nature of various exchange services. Brief comments on each of these themes are noted here.

1) **Competition between Exchanges and Other Types of Trading Systems:** Such competition keeps the services which exchanges and trading systems offer at a low price, including transaction, clearing and listing services, and provides an incentive for them to innovate and enhance their product lines, for example by automating various aspects of their systems and trading different assets.

There is a powerful reason why trading in any particular security is likely to be concentrated on a single dealing mechanism, or at most a small number of competing systems. Successful trading systems benefit from a positive network externality, because the likelihood of a market participant receiving an execution of his order on a trading system is higher if other participants also send their orders to the same trading system. Order flow thus attracts further order flow. It is therefore very difficult for new trading systems to be successful, and many new trading systems fail.

If, however, trading is fragmented between different trading venues for whatever reason, it has been argued theoretically by Mendelson (6/1987) that this dispersal of orders lowers the probability of execution at each location and may therefore reduce liquidity. He contrasts the performance of several consolidated and fragmented market architectures using as performance measures the overall gains from trade, the price volatility faced by traders, the quantity traded and the noise in transaction prices. He argues that fragmentation increases the volatility of transaction prices, reduces the overall gains from trade as well as the quantity traded, and shows that fragmentation has worse effects for thinner markets. Other theoretical analysis of the competition between exchanges identifies the possibility that different trading systems may co-exist, if traders are heterogeneous, and if each trading system satisfies the trading preferences of a particular group of investors [see, for example, Pagano (1989) and Glosten (1994)]. Some theoreticians show that competition between trading systems is not always beneficial, and that it can sometimes decrease, as well as increase, liquidity [Parlour & Seppi (10/12/2001)].

In the European context, a range of empirical analyses have been made of the competition between SEAQ International on the London Stock Exchange (LSE) and various continental European exchanges following the Big Bang in 1986 until the mid-1990s [Anderson & Tychon (4/1993), Demarchi & Foucault (2/1998), Di Noia (4/1/1998), Friederich & Tonks (3/7/2001), Jacquillat & Gresse (6/1995), Pagano (1998), Pagano, & Roëll (10/1990a), (10/1990b) and (7/1991), Roëll (1992), and Steil et al. (1996)]. This period was characterised by a bout of intense
innovation at the various exchanges concerning many aspects of their operations - their market structure, their technology, and the services they offered. Most of the measures of market performance also show that such competition was beneficial. While the analyses mostly confirm that the LSE did attract trading volume in domestic continental European securities initially, in many instances trading volume in the domestic markets did not fall. Frequently the quoted bid-ask spreads on the continental exchanges were narrower than those obtaining in London, however, often the amount it was possible to trade at these prices compared to what was possible at the prices available in London was smaller. The prices on London and on the continental exchanges appeared to be closely integrated, even though trading on London was relatively non-transparent. London offered the possibility of trading immediately against market-maker quotes, while the continental exchanges facilitated a more patient type of trading with smaller execution costs, but greater risk of non-execution.

A large amount of empirical research on the effects of competition between trading systems has also been undertaken in the US. Much of the early work concludes that increased competition in securities markets leads to tighter bid-ask spreads [see, for example, Cohen & Conroy (4/1990), Demsetz (1968), and Tinic (1972)]. These articles compare trading costs across securities and trading systems. A range of recent evidence also shows the benefits of competition between trading systems. For example, Mayhew (4/2002) examines the effects of competition and market structure on equity option bid-ask spreads between 1986-1997. Options listed on multiple exchanges have lower quoted and effective bid-ask spreads than those listed on a single exchange. Options spreads also became wider when a competing exchange de-listed an option. [See also Battalio, Greene & Jennings (10/1997), Battalio, Jennings & Selway (2/2000), Peterson & Sirri (2002) and SEC (15/4/1997) - as discussed at greater length below].

Not all the empirical evidence, however, suggests that competition between trading systems is beneficial. Porter & Thatcher (Spring 1998), for example, present evidence that increases in trading volume off the NYSE are associated with widening spreads on the NYSE. They conclude that multiple markets may be a less effective form of competition than competition via limit orders, and that a national system of limit orders may narrow spreads where the specialist’s spread is not currently dominated by limit orders [See, in addition, Davis & Lightfoot (4/1998) and Amihud, Lauterbach & Mendelson (4/2002)].

The advent of new types of trading systems has had a range of mostly beneficial effects on existing markets. Quotes on ECNs are narrower than dealer quotes [Huang (6/2002), Weston (7/2001)], and are also posted more rapidly than dealer quotes [Huang (6/2002)]. The quoted depth on ECNs is greater than through competing dealers [Weston (7/2001)]. Most studies show that more private information is revealed through ECN trades, via permanent price impacts, than through
competing dealer markets [Barclay, Hendershott & McCormick (18/2/2002), Conrad, Johnson & Wahal (2001), Huang (6/2002) and Weston (7/2001)]. This is probably because the anonymity of many ECNs allows informed investors to hide their identities. Any adverse selection effects that could lead to wider spreads on ECNs due to an increased amount of informed trading through such venues appear, however, to be outweighed by the overall lower transactions costs.

2) **Different Preferences of Different Traders:** The existence of different trading systems with distinct market architectures allows the preferences of different types of traders to be satisfied. Evidence of the diversity of types of investors is provided by the fact that investors choose to use different trading venues, and also in survey evidence. Amongst the many important differences between traders are whether they trade in large or small size, whether they are informed or uninformed, and whether they are patient or impatient.

Different benefits are available from different types of trading systems. The transaction costs of demanding immediacy of execution can be significantly higher than the costs incurred in trading patiently [Keim & Madhavan (1997)]. Large cost savings can also be achieved by using automated trading systems instead of traditional intermediated markets [Conrad, Johnson & Wahal (2001), Domowitz (4/2001), and Domowitz & Steil (1999) and (2001)]. Direct access does away with the need for investors to pay for intermediation, and allows investors to assess how much they value the benefits provided by financial intermediaries. Customer-to-customer interactions generate much lower trading costs for large trades when there is a natural counterparty offering significant liquidity on a crossing network [Barclay, Hendershott & McCormick (18/2/2002)]. However, because large depth is seldom available on an ECN, traders seeking to deal in large size typically require the services of an intermediary.

The existence of multiple trading venues, and the diversity amongst the preferences of investors, have put pressure on the notion of best execution. Contrary to the historically accepted approach, it has been suggested both that the concept of a best price is not well defined in the context of multiple trading systems [Blume (2002)], and that attempts to mandate best execution as a consumer-protection device can limit competition between trading systems [Macey & O’Hara (7/1997)].

3) **Inter-Dealer Competition:** Such competition keeps the services which dealers offer at a low price, provides an incentive for them to enhance their products, and can also reduce the market power of price-setting agents and thus result in better execution conditions. One example of empirical evidence supporting this is provided by Boehmer & Boehmer (11/6/2002), who examine the entry of the NYSE to compete for the trading of some Exchange Traded Funds (ETFs) listed and traded on the American Stock Exchange (AMEX). They find that the entrance of the NYSE leads to large declines in quoted, effective, and realised spreads, a substantial increase in quoted
depth across all market centres, and that the competition for order flow does not seem to affect price discovery adversely. Boehmer & Boehmer conclude that this improvement in market quality is not consistent with a competitive market before the NYSE entry, and that ETF market makers on AMEX earned significant monopolistic rents before the NYSE participated in that market.

4) Price Priority: With price priority, the highest bids and lowest offers get executed before other orders. A lack of price priority may lead to a reduction in the informational efficiency and quality of price formation on a market. Although it may be more difficult to deliver price priority in a market with competing trading systems than in one centralised trading system, for example when block transactions trade in one market through prices in another market, market participants themselves have an incentive to ensure that they receive the best price, and thus that price priority obtains. The mechanism by which this is delivered is that of arbitrage. If ever there are differing prices on different trading systems, then it would be profitable for a market participant to buy from the cheaper venue and sell at the more expensive venue. This will continue until the prices on both systems are equalised, or at least until the difference is no more than the cost of dealing on both systems.

Much early empirical work concludes that the NYSE offers better prices than those available on other trading venues [Angel (1994), Bessembinder and Kaufman (12/1997), Blume and Goldstein (1992) & (3/1997), Hasbrouck (1995), Huang and Stoll (7/1996), Lee (7/1993), Petersen and Fialkowski (4/1994), and SEC (15/4/1997)]. This evidence has been used to suggest that trades executed off-exchange (i.e. off the NYSE) may therefore not obtain best execution, in terms of receiving the best available execution prices.

In addition, it has been widely argued that payment for order flow diverts uninformed orders from the primary trading venue, and leaves it only the more informed orders for execution [See Harris (1995), Chordia & Subrahmanyam (1995), Lin, Sanger & Booth (1995), Easley, Kiefer & O’Hara (7/1996), Battalio (2/1997)]. This “cream-skimming” may undermine the process of price discovery because the diverted orders are cleared at prices set by the primary market and do not contribute to the price discovery there. If cream-skimming occurs, then matching the quote from the primary exchange for these diverted orders may also raise execution costs for all traders, since trading prices are based on an order flow that is now substantially riskier than it would have been in the absence of the diversion. Internalisation also weakens incentives to quote aggressively, because having the best quote in a market does not guarantee that a firm will receive the order flow, if other market participants can match its quotes, and purchase order flow.

As discussed in greater detail below, however, there is growing debate about the merits of the above arguments. Recent evidence from the literature on internalisation and preferencing suggests that market quality is not any worse when trading off the primary market (i.e. the NYSE in
the US), and furthermore that the quality of such execution, when factors other than price are taken account of, may indeed be better.

As Harris (2002), Schwartz (1991), and Stoll (6/5/2002) note, a key difficulty in assessing the costs and benefits of fragmentation is that competition between orders works best in a consolidated market, whereas competition between exchanges, trading systems, and OTC trading, and all the benefits it yields, implies fragmented markets. The two forms of competition may be incompatible with each other.

5) **Secondary Priorities:** The standard secondary priority is that of time priority. Such a rule gives market participants an incentive to submit limit orders early to a trading system, because by doing so they are more likely to have their orders executed. Time priority, and indeed any other secondary priority, is not, however, important for market orders, and is difficult to enforce in a world of competing trading systems.

Cohen, Conroy, & Maier (1986) examine theoretically the effects of not having a secondary priority rule in fragmented markets. They argue that if this occurs, even when price priority obtains over all markets, there is less chance that an order will find an available counterpart than in a concentrated market. This implies that bid-ask spreads may rise, and that the probability of a limit order being executed may fall.

Advances in automation may, however, decrease the importance of having an enforceable secondary time-priority rule. In a world in which real-time information is relatively cheaply available, and in which it is possible both to submit and withdraw orders relatively easily to, and from, a trading system, the value to an investor of secondary time priority decreases. This is because a similar outcome can be achieved by monitoring the market closely and submitting a market order should the price of the relevant stock move to where the limit would previously have been placed. This strategy also minimises the costs to investors of providing free options in the submission of their limit orders.\(^\text{14}\)

Harris (forthcoming 2002) also notes that the requirement of time priority across markets guarantees that a limit order with time priority will be executed, even if it is placed in a small illiquid market, so long as the limit price is reached and the order has time priority. A time priority rule therefore creates incentives to form new markets.

6) **Consolidation of Information:** The consolidation of information requires that all quote and trade information about the trading in a particular security be readily available from a single source. It is sometimes argued that the consolidation of such information is easier in a consolidated market. Private solutions to consolidating such information across markets have also, however, been developed in many contexts.
Transparency is widely thought to play a key role in consolidating markets, by ensuring that prices are equalised over all trading venues. Harris (forthcoming 2002) identifies three mechanisms that together consolidate a fragmented market. First, traders on one trading system must adjust their orders to reflect information that traders reveal on other trading systems. Second, some traders must route their orders to trading systems where they expect to obtain the best prices. Finally, arbitrageurs must trade whenever prices on one trading system are inconsistent with prices on another trading system. Harris argues that these three mechanisms will consolidate a fragmented market only if information about trades and orders in each market is publicly available at low cost. Without this information, traders will not be easily able to search for the best price across markets.

This view is not universally accepted, however. It may be that as long as some arbitrage exists between trading systems, then markets can be consolidated, in the sense of having equal prices without full transparency. One theoretical model which may confirm this notes that all that may be required to effect consolidation of markets is that the brokers operating in the market have a sufficient number of interactions. Galetovic & Zurita (3/2002) suggest that when the value of liquidity stems from the ability to reach potentially as many traders as possible, the market is integrated when every broker meets every other broker in at least one exchange. Thus, fragmentation is not about trades being executed on different exchanges, but about connections among brokers.

The costs and benefits of pre- and post-trade transparency are discussed at greater length in the final section of the report.

7) **Public Good Nature of Various Exchange Services:** Such services include price stabilisation, market surveillance, self-regulation, and quality certification. The existence of multiple centres for trading an asset may reduce the provision of the public goods that exchanges provide, given the difficulty of funding them. It may also be easier to share regulatory costs on a fair basis in a consolidated market. However, various regulatory solutions have been proposed and implemented to address the problems of ensuring that trading on all available venues is adequately regulated, and that such regulation is funded appropriately.

2.2. **ATS, ECNs, & Crossing Networks**

There is a growing literature about the effects of new types of trading systems (including ATSs, ECNs and crossing networks) on the performance of more established markets. A relatively detailed and technical summary of the key theoretical and empirical literature about such trading systems is presented in this part. Before summarising some relevant articles on the topic, however, a few introductory comments are first made.

Definitions of the new types of trading systems (i.e. ATSs, ECNs and crossing networks) are
not provided here, given the many different ways in which the concepts are used, and the many differences between such trading systems. Some typical attributes of the different types of systems are, however, notable. They frequently employ electronic trading systems. Some allow investors to deal directly with each other, without requiring any intermediation. They employ different levels of transparency. Many allow anonymous trading. Typically ATSs and ECNs allow participants to submit limit orders, and thereby undertake price discovery, while crossing networks use prices determined elsewhere to set their execution prices, and thus do not contribute to price discovery. The different types of systems frequently seek different types of clients, for example retail or institutional investors.

Domowitz (4/2001) and Domowitz & Steil (1999) and (2001) show that large cost savings can be achieved by using automated trading systems over traditional intermediated markets. They also confirm that automated trading venues are not cheaper only because “easier” trades are more often sent to them, by showing that savings from automated execution are present even for more difficult trades. Direct access does away with the need for investors to pay for intermediation, via brokers and dealers, and thus allows investors to assess how much they value the benefits provided by financial intermediaries.

Huang (6/2002) compares the quality of quotes submitted by ECNs and traditional market makers to the NASDAQ quote montage for the most active Nasdaq stocks. He shows that ECNs not only post informative quotes, but also that their quotes are narrower and posted more rapidly than dealer quotes.

Barclay, Hendershott & McCormick (18/2/2002) examine the differences between ECNs and Nasdaq market makers. The ECNs’ customer-to-customer interactions generate much lower trading costs for large trades when there is a natural counterparty offering significant liquidity on an ECN. However, because large depth is seldom available on an ECN, traders seeking to deal in large size typically require the services of an intermediary. More private information is generally revealed through ECN trades, via permanent price impacts, than through market-maker trades, even though more trades occur with market makers.

Weston (7/2001) examines the growth of trading on ECNs. He finds that ECNs compete directly with traditional Nasdaq dealers, forcing lower average quoted, effective, and realised spreads over all dealers, and that ECNs improve the average quoted depth. He also finds, however, that trading on anonymous ECNs increases the adverse selection component of the spread, thus making prices less informationally efficient. Chen, Chou & Shyy (undated) show that the spreads are much smaller on ECNs than in the dealer market, and that spreads decrease as the participation rate on ECNs increases.

Dönges & Heinemann (25/1/2001) and Hendershott and Mendelson (10/2000) analyse
theoretically competition between dealer markets and a crossing network. Dönges & Heinemann (25/1/2001) show that under certain circumstances, assets with low volatility and large volumes are traded on crossing networks, while others are crossed on dealer markets. Efficiency, however, requires more assets to be traded on crossing networks.

Hendershott and Mendelson (10/2000) find that a crossing network has two countervailing effects on a dealer market. The first is a positive “liquidity” externality, by which ECN trading increases the probability that a given trade will be executed on the low-cost alternative. Traders who use the crossing network exclusively can provide a counterbalancing effect that reduces adverse selection and inventory costs for market makers on the central market, which in turn lowers costs on all trades. The second is a negative “crowding” externality, by which low-liquidity preference trades compete for execution on the ECN with high-liquidity preference trades and may “crowd out” the crossing of high liquidity preference trades. When traders use the dealer market as a “market of last resort”, by submitting orders to the crossing network first, and then going to the dealer market if their orders are not executed, the dealer market becomes more risky, forcing dealers to protect themselves by widening the bid-ask spread. This can, however, lead to more efficient subsequent prices on the dealer market. Overall, the introduction of a crossing network can either increase or decrease social welfare, depending on security characteristics and trader preferences.

Conrad, Johnson & Wahal (2001) examine the execution costs of trades sent to traditional trading systems, crossing networks, and ATSs. They find that institutional orders receive lower execution costs on ECNs relative to traditional dealers, though they are typically in smaller size and have lower fill rates. ECNs facilitate more informed trades, but market makers can better execute larger trades.

Degryse & Van Achter (2001) review the impact of ATSs on liquidity in traditional market places, and distinguish two important trends. First, ATS are currently more successful in the US than in Europe. Second, within the US, there exists a divergence between the impact of ATS on the Nasdaq dealer market and on the NYSE. ATS are attracting about 30% of market share in the Nasdaq market, whereas their impact on the NYSE is rather small. They propose two reasons for the relatively greater success of ATS in the US than in Europe. First, traditional European exchanges became automated earlier than their American counterparts. Second, European markets are mostly organised as auction markets - a market structure similar to most ECNs - and thus the ECNs offer little that is new.

Board & Wells (2000) compare trading on SETS on the London Stock Exchange and Tradepoint (at that stage an ECN) concerning liquidity and best execution of UK shares. Their analysis indicates that while SETS was more active during the period under consideration,
Tradepoint did offer better prices for significant periods during the day at volumes that were roughly comparable to those offered by SETS. Offering such better prices was not sufficient, however, to attract significant business.

Gresse (5/2002) investigates the impact of crossing network-trading on the liquidity of a dealer market by comparing trade and quote data from the quote-driven segment of the LSE and internal data from the POSIT crossing network. The crossing network trading does not worsen adverse selection on the central market, and does not make the after-crosses market riskier. To the contrary, the crossing network tends to reduce the oligopolistic power of market makers on the central market, and to give them a risk sharing opportunity that leads them to improve quotes, and thus facilitates the discovery of prices. Short-term excess volatility and temporary market impact are also reduced with crossing network-trading.

3. Internalisation

The economic evidence about the effects of internalisation on market performance and best execution is summarised in this section. It is divided into three parts. In the first, a broad overview of the key themes in the literature is presented. In the next part, a more detailed and technical summary of the key theoretical and empirical literature on internalisation is presented. In the last part, the available evidence on upstairs trading, an important aspect of internalisation, is described. Given that internalisation is a form of fragmentation, many of the themes discussed in the previous section about fragmentation are repeated here.

3.1. Key Themes

The key themes that have arisen in the literature on internalisation are noted here.

The early theoretical analysis of internalisation stressed two major possible costs associated with it. First, it was shown potentially to harm the price discovery function of a market, because it stops all orders in a particular security from competing directly against each other on a single order book [Cohen, Maier, Schwartz, & Whitcomb (3/1982)]. Such fragmentation might also lead to larger bid-ask spreads and greater price volatility.

A second key problem identified theoretically with internalisation is that it may divert uninformed trades away from a primary exchange, leaving only the more informed orders on the primary exchange [Easley, Kiefer & O’Hara (7/1996)]. This “cream skimming” may lead dealers to set wider spreads on the primary exchange in order to protect themselves from greater adverse selection. If the diverted uninformed orders are executed at prices taken from the primary market, they may also then be executed at worse prices than without preferencing or internalisation.

Two key benefits arising from internalisation have subsequently been noted. First, it might
allow alternative trading venues to compete with a primary or central market, and such competition could lead to enhanced pricing. Second, it would give market participants a greater diversity of choice as to where to execute their orders, rather than only having the single option of sending them to the central market. This might benefit particular classes of investors. For example, if orders are split according to how well informed they are, and if small investors have uninformed orders, then the execution of their trades via internalisation may let them obtain better prices and lower dealing costs than on a primary exchange [Harris (2002)]. If dealers act in a competitive environment, they will set smaller spreads for these uninformed orders than would occur in a consolidated market, given that the dealers do not have to protect themselves against adverse selection as a result of dealing with informed traders.

As noted in the section on fragmentation, much of the early empirical work on internalisation (almost exclusively about the US markets) concluded that the primary exchange (the NYSE) offers better prices than those available at other trading venues. [See for example, Angel (1994), Bessembinder and Kaufman (12/1997), Blume and Goldstein (1992) & (3/1997), Hasbrouck (9/1995), Huang and Stoll (7/1996), Lee (7/1993), and Petersen and Fialkowski (4/1994)].

A growing body of more recent empirical evidence about internalisation and preferencing, however, highlights the benefits available from internalization, and questions both whether better prices are in fact available from the NYSE, and also more generally whether better execution is available. Internalisation and preferencing have been shown not to harm the execution of market orders or limit orders, and may indeed better it, using various different measures [Battalio, Greene & Jennings (10/1997), Battalio, Jennings & Selway (2/2000), Peterson & Sirri (2002), & SEC (15/4/1997)]. For example, market orders traded via preferencing on regional exchanges tend to trade more favourably relative to the NYSE than market orders placed on non-preferencing regional exchanges [SEC (15/4/1997)]; limit orders have a greater probability of executing on regional exchanges than on the NYSE [SEC (15/4/1997)]; and internalisation seems to have little short-run effect on posted or effective bid-ask spreads [Battalio, Greene & Jennings (10/1997)].

There is conflicting data about whether internalisation leads to cost competition or an adverse selection problem [See Battalio (2/1997) versus Easley, Kiefer & O’Hara (7/1996)]. The most recent theoretical analysis of how internalisation functions takes account both of the prices that investors receive and of the commissions they pay for their executions. More complicated conclusions are drawn about what the effects of internalisation on best execution are [Battalio & Holden (1/2001)]. In particular, it is shown that payment for order flow can lead both to an increase in execution quality, taking into account both the price and commissions paid, which is consistent with cost competition, while at the same time giving rise to an increase in the proportion of trading
on a primary exchange which is informed, which is consistent with cream-skimming.

There is empirical evidence that when best execution takes account of factors other than merely price, the internalisation of orders can lead to better executions than are available on the NYSE. For example, while execution on the NYSE appears better than via third-market firms using measures of trade-price quality, third-market dealers provide more timely executions and produce more liquidity enhancement than at the NYSE [Battalio, Hatch & Jennings (3/2000)]. If potential reductions in commissions are taken into account as well as execution prices, previous results showing that better execution has been available on the NYSE than elsewhere may need to be revised [Battalio & Holden (1/2001)]. Previous results showed that of 13 comparisons of execution quality on the NYSE versus at non-NYSE venues (as measured solely by execution prices), 11 favoured the NYSE. If account is taken of possible rebates in commission income, however, the majority of the comparisons (7 out of 13) could favour the non-NYSE venues over the NYSE in terms of combined cost of trading.

A range of theoretical reasons have been put forward for why institutional traders may like to use the upstairs market, a key form of internalisation. The ‘upstairs’ market is a US term referring to the activities of members of an exchange, namely brokers and dealers, when they search for counter-parties for big institutional orders off the floor of the exchange. Upstairs markets may help investors locate trading counter-parties, and thereby execute large trades without fully revealing their orders to the downstairs market [Burdett and O’Hara (1987) and Seppi (3/1990)]. Upstairs counter-parties may be able to filter out large orders containing adverse information. Upstairs trading may facilitate the collection of information about the unexpressed supply of, and demand for, securities [Grossman (1992)]. Upstairs trading may help risk-sharing amongst market intermediaries, thereby lowering transaction costs [Keim & Madhavan (1996)].

Empirical evidence from a range of different contexts and jurisdictions shows the benefits of upstairs trading. On the NYSE, upstairs trading appears to be used by investors who can signal credibly that their trades are liquidity motivated [Madhavan & Cheng (1/1997)]. The upstairs market may thus enable transactions that would otherwise not occur in the downstairs market. In Australia, off-exchange trading - including ECN, upstairs and after-hours trading - is shown to benefit those traders in a position to switch trading venues, by lowering their trading costs [Fong, Madhavan & Swan (2/3/2001)]. In Canada, the upstairs market makers on the Toronto Stock Exchange provide a vehicle for screening out informationally-motivated orders and for executing large liquidity-motivated orders at a lower cost than the downstairs market [Smith, Turnbull & White (10/2001)].
3.2. A More Detailed Summary

A relatively detailed and technical summary of the key theoretical and empirical literature on internalisation is presented in this section. Relevant articles are outlined in chronological order.¹⁶

Cohen, Maier, Schwartz, & Whitcomb (3/1982) argue theoretically that brokerage houses and their customers have private incentives which favour in-house execution, but that when this occurs the market as a whole will suffer. With such internalisation, the fragmented market has larger bid-ask spreads and greater price volatility than without. They conclude that the best market structure is an informationally-consolidated market where the best in-house quotes are exposed to the entire market and secondary (time- or size-based) priority rules are also enforced on a market-wide basis.

Easley, Kiefer & O’Hara (7/1996) test whether the purchase of order flow by brokers “cream-skims” the uninformed liquidity trades, leaving the information-motivated trades to be executed more expensively on the established markets. Their main empirical result is that there is a significant difference in the information content of orders executed on the NYSE and on the Cincinnati Stock Exchange (CSE - where purchase of order flow occurs), with those on the NYSE conveying more information, and that this difference is consistent with “cream-skimming”. Since the diverted orders are less risky, an adverse selection problem arises with respect to the remaining order flow. Easley, Kiefer & O’Hara suggest that the more successful the policy of segmentation is, the more profitable it will be for the purchasers of order flow. If they agree to match the prices on the NYSE, but these prices widen because they reflect a high adverse selection problem for dealers on the NYSE, then purchasers of order flow will be buying orders with a low adverse selection risk, but taking the profit (i.e. the bid-ask spread) which is set to take account of a high-risk order composition. Easley, Kiefer & O’Hara note that while their results support the existence of cream-skimming, they cannot reject the possibility of simultaneous profit sharing, whereby the purchasers of order flow may also be sharing their profits with end investors.

Battalio (2/1997) compares the bid-ask spread for the NYSE-listed securities before and after a major third market broker-dealer (Madoff) begins to purchase selectively and execute orders in those securities. Tests reveal that the time-weighted quoted bid-ask spread tightens when Madoff enters the market, and trading costs, as measured by the difference between the transaction price and the mid-point of the contemporaneous bid-ask spread, do not widen. Together these results suggest that the adverse selection problem associated with allowing agents to execute orders selectively in exchange-listed securities may be insignificant. Battalio does note, however, that the results arising from Madoff’s activity may not be generalisable to other brokers.

The SEC (15/4/1997) undertook some empirical analysis on preferencing, by analysing trading activity on the NYSE (the primary market), the CSE and the BSE (regional exchanges with
formal preferencing programs), and on the CHX, PSE, and PHLX (regional exchanges without formal preferencing programs). The SEC concluded that its data did not support the view that preferencing has harmed the execution of market orders or limit orders. Market orders traded on preferencing regional exchanges tend to trade more favourably relative to the NYSE than market orders placed on non-preferencing regional exchanges. In addition, limit orders have a greater probability of executing on regional exchanges than on the NYSE. Further, the data do not indicate that preferencing has harmed the market quality of the preferencing exchanges. Indeed, the CSE often provides competitive markets. However, the SEC noted that its results do not conclusively demonstrate that preferencing has improved the marketplace because it was not possible to observe directly the effect of preferencing on quote-based competition.

Battalio, Greene & Jennings (10/1997) show empirically that the opportunities the BSE and the CSE provide for their members to take the other side of their customers' orders through affiliated market-makers, namely by internalisation, have little short run effect on posted or effective bid-ask spreads. This is despite substantial movement of order flow away from the NYSE when trading under one of these regional stock exchange programs begins. They conclude that there is no significant evidence of worsening efficiency on market quality due to regional fragmentation of the NYSE. Battalio, Greene & Jennings do, however, identify a series of caveats associated with their findings.

Chakravarty & Sarkar (4/1998) study a range of issues related to brokers’ trading in a theoretical model in which many informed traders and noise traders trade through many brokers. Brokers may trade with their customers in the same transaction (simultaneous dual trading) or trade after their customers in a separate transaction (consecutive dual trading). The simultaneous dual trading model, in which brokers act as market makers to noise traders by offsetting a portion of their order flow in the same transaction, captures the essence of internalisation of uninformed order flow. In one version of the model, Chakravarty & Sarkar assume that a fixed subset of brokers are involved exclusively with internalising the order flow of noise traders, while the remaining brokers are involved exclusively with mimicking informed traders. With such internalisation, brokers’ commissions are lower, and the price, net of commissions, may also be lower. In another version of the model, Chakravarty & Sarkar allow brokers free entry into internalisation activity, and find that although internalisation results in lower brokers’ fees, both market depth and price informativeness are also lower, and uninformed losses are higher. They also show that internalisation is likely to be more prevalent in thin markets with few informed traders.

Battalio, Jennings & Selway (2/2000) study the division of market-making revenue among dealer, broker, and trader. They find that many orders routed based on order-flow inducements enjoy lower net trading costs, taking into account trading price net of commissions. Their findings
are consistent with the claim that brokers accepting order-flow payments share these revenues with traders through lower commissions. They argue that this evidence suggests that trading with brokers taking payment for order flow is not unambiguously harmful to investors, relative to trading through brokers not accepting such order-routing inducements, and that conclusions regarding across-market execution quality based solely on transaction prices may be misleading. Battalio, Jennings & Selway do note, however, a series of cautions about whether their findings are generally valid, including the possibility that the existence of order-flow payments could increase the cost of liquidity for all traders.

Battalio, Hatch & Jennings (3/2000) assess market-order execution quality as a multidimensional concept. They compare execution quality on the NYSE and at a third-market dealer, by evaluating execution speed, trade price, and the ability to trade more than the quoted size at the quoted price. They find that each trading venue provides high-quality executions on different dimensions of quality. On average, the NYSE betters the third-market firm using the traditional measures of trade-price quality, producing higher price improvement rates and lower liquidity premia. The third-market dealer, however, provides more timely executions and produces more liquidity enhancement than the NYSE. These results suggest that no single venue may provide best execution for all retail market orders, implying that the broker’s routing decision is complex. With an overall execution-quality measure for market orders, they find that the two trading venues offer very similar execution quality. If payment for order flow is considered, then it appears that the third-market firms earn less revenue providing liquidity to market-order traders than does the NYSE. This suggests that payment for order flow and high-quality trade executions are not necessarily mutually exclusive.

The complexity of internalisation is highlighted by some theoretical analysis presented by Battalio & Holden (1/2001). They show that externally-verifiable characteristics of traders and/or orders (such as for example whether a trader is a professional or not) allow profitable order purchasing or profitable internalisation, even when the permitted differences between prices are infinitesimally small and when primary dealers are competitive. This is because such externally-verifiable characteristics may permit a sorting of orders based on the likelihood of information content. Battalio & Holden define a new measure of the cost of trading called Total Trading Cost (TTC) to measure execution quality. TTC is defined as the effective half spread (EHS) plus the broker’s per share commission. The EHS is the absolute value of the transaction price minus the quote midpoint. TTC is the entire cost of trading faced by an outside trader. It is important to include the broker’s dollar amount per share commission when calculating this entire cost of trading, because there are differences across equilibria in the size of the payment for order flow and in what gets passed on to the uninformed trader.
Battalio & Holden use the concept of TTC to reconcile the apparently contradictory results of Battalio (2/1997) and Easley, Kiefer & O’Hara (7/1996). Battalio (1997) shows that the TTC drops when a major broker begins purchasing order flow, which he interprets as consistent with cost competition, while Easley, Kiefer & O’Hara (7/1996) find that the probability of informed trading is lower on the CSE than on the NYSE, and interpret this evidence to be consistent with cream-skimming. Battalio & Holden show that it is possible for the TTC to fall when purchasers of order flow enter the market, while at the same time the nature of the order flow that is purchased, and thus diverted away from the primary market, can be less informed than the orders that are still sent to the primary market. This means that internalisation can deliver both cost competition, and cream-skimming.

Battalio & Holden also re-examine the empirical literature on trading costs on, and away from, the NYSE. They study 13 comparisons from five different articles, all of which use the EHS as the basis of comparison. 11 of these 13 comparisons favour the NYSE and only 2 favour the non-NYSE venue. They assess, however, the extent to which, the comparisons would continue to favour the NYSE if part of the payment for order flow were returned to investors via reduced commissions. They find that of the 11 comparisons that initially favoured the NYSE, 9 might be reversed depending on the degree of competition of the brokerage industry. For example, if the brokerage market is sufficiently competitive that brokerage firms are forced to pass back over 50% of the payment for order flow, then 5 of the remaining 11 comparisons would be reversed. This would mean that the majority of the comparisons (7 out of 13) would then favour the non-NYSE venue over the NYSE in terms of combined cost of trading and commissions paid.

Parlour & Rajan (4/11/2002) examine the practice of payment for order flow by developing a theoretical model of broker and dealer markets. Competing market makers quote bid-ask spreads, and competing brokers choose commissions to be paid by an investor. Investors, who can submit either market or limit orders, choose a broker to minimise transaction costs. With no payment for order flow, Parlour & Rajan find there exist equilibria in which brokers and market makers earn positive profits. There is also an equilibrium in which they earn zero profits. With payment for order flow, however, spreads more widen than is needed to compensate for this payment. Consumer and social welfare are both lower in any equilibrium with payment for order flow, compared to the zero transactions costs equilibrium. Payment for order flow also redistributes payoffs from traders who demand liquidity to those who supply it, and decreases overall welfare.

Peterson & Sirri (2002) analyse market quality on the NYSE (the primary market), the CSE and the BSE (regional exchanges with formal preferencing programs), and on the CHX, PSE, and PHLX (regional exchanges without formal preferencing programs). Their analysis rebuts the hypothesis that preferencing harms the market by leading to worse executions on all exchanges.
While they cannot directly observe the effect of preferencing on all exchanges, their data indicate that the effective spreads of market orders trading on preferencing regional exchanges tends to be lower than the effective spreads of market orders trading on non-preferencing regional exchanges. In addition, limit orders have a greater probability of executing on preferencing regional exchanges than on the non-preferencing regional exchanges. In most cases the market quality of preferencing regional exchanges is superior to the market quality of non-preferencing regional exchanges.

3.3. Upstairs Trading

Upstairs trading, whereby a member of an exchange may seek to cross a large trade off the order book of the exchange, is an important type of internalisation. A range of reasons have been put forward for why institutional traders may like to use the upstairs market. Burdett and O’Hara (1987) and Seppi (3/1990) suggest that upstairs markets help investors locate trading counterparties, and thereby execute large trades without fully revealing their orders to the downstairs market. Upstairs counter-parties can filter away large orders containing adverse information. Grossman (1992) argues that upstairs trading may facilitate the collection of information about the unexpressed supply of, and demand for, securities. Keim & Madhavan (1996) argue that upstairs trading can help risk-sharing amongst market intermediaries, thereby lowering transaction costs.

Madhavan & Cheng (1/1997) analyse the ability of various market mechanisms to provide liquidity for large equity trades. They find that the downstairs NYSE floor market is a significant source of liquidity. Although negotiation in the upstairs market provides better execution than the downstairs market for large trades, these differences are economically small. They find, however, that upstairs markets are used by traders who can credibly signal that their trades are liquidity motivated. Liquidity providers, especially institutional traders, are reluctant to submit large limit orders, and thus offer free options to the market. Upstairs markets allow these traders to participate selectively in trades screened by block traders who avoid trades that may originate from traders with private information. Madhavan & Cheng suggest that the upstairs market’s major role may thus be to enable transactions that would otherwise not occur in the downstairs market.

Fong, Madhavan & Swan (2/3/2001) analyse the nature and determinants of order flow fragmentation both over time and across all securities listed on the Australian Stock Exchange, and find the off-market volume to be strongly related to various measures of liquidity on the primary market. Off-market trading is seen as an alternative source of liquidity to supplement the primary market. Fong, Madhavan & Swan conclude that there is significant competition between markets in highly liquid securities, and also that off-market trading benefits those traders in a position to switch trading venues by lowering their trading costs.
Smith, Turnbull & White (10/2001) find that the wider the bid-ask spread on the Toronto Stock Exchange (TSE), and the lower the depth on the opposite side of the limit order book, the more likely trades are to be executed upstairs. They test the hypothesis that upstairs intermediation lowers adverse selection cost on the TSE, and find that upstairs market makers do effectively screen out information-motivated orders and execute large liquidity-motivated orders at a lower cost than the downstairs market. In one-quarter of the trades, the upstairs market offers price improvement over the limit orders available in the consolidated limit order book. Smith, Turnbull & White conclude that upstairs markets do not cannibalise or free ride off the downstairs market.

4. Transparency

The economic evidence about the effects of transparency on market performance and best execution is summarised in this section. It is divided into four parts. In the first, a broad overview of the key themes in the literature is presented. In the next three parts, more detailed and technical summaries of the literature on pre-trade (or quote) transparency, on post-trade transparency, and on the effects of transparency on market linkages, are presented in turn.

4.1. Key Themes

The key effects of transparency on securities markets, and on the participants in such markets, are complex and contradictory.

Almost all the evidence shows that greater transparency improves informational efficiency. It may allow traders to select which trading system delivers the best quoted price, thus facilitating arbitrage between different systems, ensuring price priority, and enhancing the price discovery process. In turn, this is often believed to enhance best execution.

There is growing evidence, however, that greater transparency may also harm market performance in various ways. Not all investors are willing to expose their orders publicly, given the free option they provide by doing so to potential counter-parties. Enhanced transparency can reduce their willingness to participate in the market, and has been shown to decrease liquidity in various contexts.

Informed traders are likely to prefer anonymous trading systems, so that they can hide their trading intentions, while uninformed traders prefer greater transparency. No single transparency regime will therefore be seen as optimal by everybody.

Transparency may make bid-ask spreads widen because market makers have less incentive to pay to capture the information that a trade with an informed trader will bring. Alternatively spreads may decrease because information about transactions reaches all market participants, or because dealers are aware of each others’ positions and compete more strongly with each other.
Transparency may encourage stabilising speculation that helps absorb order flow imbalances and reduce volatility. Alternatively it may exacerbate market participants’ strategic behaviour towards each other, with the possibility of increasing volatility.

The level of transparency in a market affects how trading systems compete against each other. Several commentators have suggested that exchanges and trading systems should own the property rights in the information arising from their trading systems, and that they have the appropriate incentives in most circumstances to determine the appropriate levels of transparency for their trading systems.

4.2. Quotes

A relatively detailed summary of the literature examining the effects of pre-trade or quote transparency on market performance and best execution is presented here.\(^{20}\) The literature is described in chronological order. The articles span a range of different market structures, jurisdictions and contexts, and also employ all three of the analytical techniques of theory, experimentation, and empirical analysis.

The effects of improving the means of quote and trade dissemination in a historical context have been investigated by Garbade & Silber (1978). They study the consequences of three advances in information technology on the integration of various geographically separated markets: the establishment of the US domestic telegraph system during the 1840s, the opening of the trans-Atlantic cable in 1866, and the implementation of the Consolidated Tape Association (CTA) in 1975. Garbade & Silber find that the first two advances reduced price differentials between the affected markets, while the third change had little effect. This appeared to be both because the information on the CTA was already available, and because the CTA did not allow orders to be transmitted more quickly back to the suppliers of liquidity, unlike the first two advances. Garbade & Silber note that the first two innovations were established by the private sector, while the CTA, in contrast, was implemented as a result of a regulatory initiative. They contend that without such regulatory intervention, the consolidated tape would not have been established, because it would have been thought unnecessary given the existing relatively short time of information transfer, and the minimal economic benefits that were obtainable.

Garbade, Pomrenze & Silber (1979) investigate the extent to which the quotes of competing dealers hold information of value to each other in the US GNMA pass-through securities market. They find that although dealers acquire new information from their competitors, they do not consistently treat their own quotes as redundant information once they had observed those of their competitors. The quality of the information carried in a particular set of observed quotes depends on the dispersion of those quotes. In particular, the more compact the distribution of competing
dealers’ prices is observed to be, the more dealers revise their estimates of the equilibrium price towards the mean price of this observed distribution.

Umlauf (1990) examines the empirical effects of an informational asymmetry in the market for US Treasury securities, which arose because of the existence of two different types of dealers in the market. Primary dealers were allowed to trade through all Inter-Dealer Brokers (IDBs) operating in the market, while secondary dealers were only allowed to trade through a subset of the full group of IDBs. The secondary dealers were therefore at a disadvantage to the primary dealers, in that they did not have access to the full set of trading information relayed on all the IDBs’ screens. Umlauf finds that this informational asymmetry affected the transactional characteristics of the market in two ways: secondary dealers’ price expectations lagged those of primary dealers, and the bid-ask spreads of secondary dealers were wider than the corresponding ones of the primary dealers.

Forster & George (1992) examine theoretically the effects of liquidity traders being able to advertise whether they are net buyers or sellers, or the magnitude of their orders. The disclosure of the net direction of liquidity orders to all market participants in advance of trading decreases the expected transactions costs paid by liquidity traders. The informativeness of prices is independent of whether, and to whom, information about the net direction of liquidity trades is revealed. The disclosure to all types of market participants of the size of liquidity traders’ orders decreases their transaction costs only if there is enough competition between market participants with private information about the security’s fundamental value. It does not, however, affect the average price informativeness. Disclosure of the magnitude of liquidity trades increases the incentive for acquiring costly private information that is possessed by only a few traders, but decreases the incentive for acquiring inexpensive private information that is widely known.

Biais (1993) analyses theoretically the performance of two types of dealer markets: a “centralised” market, in which dealers compete to attract order flow, and can observe the quotes and transactions of their competitors, and a “fragmented” market in which transactions arise as a result of bilateral negotiations, and market makers cannot observe their competitors’ quotes. Dealers are assumed to be risk-averse, and thus to incur a cost when they take-on some inventory, but face no costs arising from asymmetric information as there are no informed traders in the market. The quote-setting behaviour of dealers is shown to be dependent both on their own inventories, and on the information that they have about their competitors’ quotes and inventories. In the fragmented market, market makers take advantage of the lack of transparency to post wider bid-ask spreads, and to earn monopolistic profits. While the average market bid-ask spread is equal in both market structures, it is more volatile in the centralised market than in the fragmented one.

Madhavan (7/1996) investigates theoretically the extent to which the dissemination of information about orders affects the performance of a batch auction market. Although transparency
encourages stabilising speculation which helps absorb order flow imbalances, it also exacerbates traders’ strategic behaviour towards each other. Informed traders profit from the difference between the expected and the actual price, and are aware that if they attempt to trade to take advantage of a difference between the actual and the expected price, other traders will follow suit, thereby reducing the expected difference. They may therefore reduce their order size so as to maintain a wider differential between the expected and the actual price. In extreme circumstances, this may lead to a market failure. Transparency does, however, lead to more informative prices as long as there is no market failure, and also reduces price volatility and increases liquidity as long as the market is sufficiently large and liquid.

Pagano & Roëll (1996) compare theoretically the liquidity that obtains on four different market structures: a transparent batch auction; a relatively opaque batch auction; a continuous auction in which all past orders and trades are visible; and a continuous dealer market with no publication of trades within a trading period. Each market contains an informed trader, who may follow an exogenously determined strategy, or who may condition his trading strategy on the market structure in which he is operating. The relative performance of the different markets is complex.

Trading costs for uninformed traders are lower in the transparent batch market than in the dealership market, with a given trading strategy for the informed trader. Liquidity in the opaque batch auction cannot be unequivocally compared, however, with the continuous auction or the dealer market. If the informed trader can pre-commit to a given trading strategy, the expected trading costs of liquidity traders are higher on average in the dealer market than in the transparent auction market. There may, however, be a range of trade sizes for which the dealer market provides better prices for liquidity traders. Without pre-commitment, various scenarios are possible. If all liquidity traders’ orders are of the same size, transparency unambiguously enhances market performance: the transparent batch market is more liquid than the continuous auction market, which in turn is more liquid than the dealership market. If the optimal strategy of the informed trader differs between markets, the liquidity of the auction market may, however, no longer be better for all noise traders. For example, a noise trader placing a small order may prefer the dealership market over the transparent auction. The transaction costs of liquidity traders averaged across all trade sizes, however, are lower in the transparent auction than in the dealership market.

Ui (1996) investigates theoretically the optimal level of transparency that minimises price volatility. He considers two types of transparency. One is transparency for public information where all traders commonly observe the same order flow. The other is transparency for private information where different traders observe different and independent order flows. In the case of transparency for public information, when the variance of order flow is large enough and the market is not
transparent, increasing transparency reduces price volatility. Too much transparency, however, increases price volatility. In the case of transparency for private information, when the variance of order flow is large enough, increasing transparency reduces price volatility, and the most transparent markets enjoy the least price volatility.

Lamoureux & Schnitzlein (1997) compare experimentally a transparent market where all trading is channelled centrally to a group of competing dealers, with a similar market where traders may also trade with another via an opaque process of bilateral search over a series of periods. They find that when there is no bilateral search alternative, both dealer profits and liquidity trader losses are large. When bilateral search is allowed liquidity traders do much better, however, and dealer profits are close to zero. Dealers compete more aggressively in the search alternative initially than they do against each other in markets without a search process. The profits of informed traders are, however, greatest when opaque trading is allowed. Market efficiency is not harmed by the introduction of the bilateral search process. Search trades tend to be set at prices close to those obtaining in the dealer market prior to its close. Trades in the market with the search alternative are closer to the asset’s true value during the search process than in the trading intervals before and after the halt. They are also generally closer to the asset’s true value than at any time in the market without the search alternative.

Flood, Huisman, Koedijk & Mahieu (1/1999) investigate how quote transparency affects the informational efficiency of a multiple-dealer market. Three market structures with varying amounts of quote information made public are examined: 1) a “centralised” market with live quotes from all market makers appearing continuously on a trading screen; 2) a “modified centralised” market, in which only the best bid and ask quotes from all the market makers are displayed; and 3) a “decentralised” market, with no quote information publicly disseminated, but prices and transactions are communicated on a bilateral basis with customers calling dealers, and dealers calling other dealers. In real time, there are no clear differences in the informational efficiency of the centralised and decentralised markets. However, in “transaction” time, a scaling of time in which a unit of time passes after each transaction, transaction prices converge significantly more quickly to their fundamental values in the decentralised market than in the centralised markets.

Scalia & Vacca (1999) analyse a change in the degree of pre-trade transparency on MTS, an electronic inter-dealer market for Italian Government bonds, which occurred in 7/1997 when quotes were allowed to be anonymous. They find that the decrease in transparency makes liquidity traders worse-off, while large and informed traders find it less costly to execute block trades. The move to anonymity is accompanied by an increase in market liquidity, by a reduction in volatility, and also by an increase in the speed with which information is aggregated on MTS.21
Bloomfield and O’Hara (3/2000) build a theoretical model in which low-transparency dealers are more likely to compete to set the best quotes in early rounds of trading, in order to capture more order flow. The information they learn from the order flow then allows them to quote narrower spreads than their more transparent competitors and to avoid money-losing trades. This informational advantage declines with repeated rounds of trading because low-transparency dealers reveal their information through their choices of quotes. Moreover, as trade progresses and individual dealers learn from trade outcomes, spreads for all dealers decline. Bloomfield and O’Hara provide experimental evidence that supports all of these predictions.

In a second experiment, Bloomfield & O’Hara find that if dealers are allowed to choose whether to be transparent or not, they have a strong preference for being opaque. Low transparency dealers do not, however, drive out their more transparent competitors. There are advantages to being the only transparent dealer in an otherwise opaque market. Some informed traders may intentionally engage in loss-making visible trades in order to mislead other market participants. In addition, when there are many low-visibility dealers, no one of them has a significant informational advantage over the high-visibility dealer, as the informed trader may spread his trades more evenly among the competing low-visibility dealers. This reduces the losses that high-visibility traders incur as a result of adverse selection.

Anand & Weaver (25/9/2001) study the value limit-order traders place on the ability to hide their intentions in an open electronic limit order book, or from the viewpoint of the market-order traders, the ability to get a complete picture of the order flow available. They examine the situation at the TSE when it stopped the use of hidden limit orders for stocks traded on its electronic CATS system on 18/3/1996. At that time, traders in stocks listed on the CATS could no longer hide additional quantity behind their orders. Liquidity, market share and volume remained the same, but traders who previously had been able to hide large quotes in the book had to switch to use market orders.

Oehler & Unser (27/9/2001) explore the effects of transparency on the price formation process employing 16 experimental asset markets. A key finding is that more quote transparency, via an open order-book, lowers market depth and liquidity. New information arrival or high information intensity also causes higher volatility and trading volume, and such volatility is greater with an open order book.

Madhavan, Porter & Weaver (15/10/2001) examine the effects on the TSE of an increase in pre-trade transparency, when it publicly disseminated the limit order book on both the traditional floor and on its automated trading system. They find this led to significant increases in execution costs and volatility. These effects were, however, concentrated in floor stocks where transparency was low, and not in CATS stocks which already featured a high degree of information disclosure.
The reduction in liquidity was also associated with significant declines in stock prices, but the increase in transparency did not affect the cross-border market shares of the US and Canadian exchanges which competed for trading in the relevant stocks.

Naik & Yadav (15/3/2002) examine a series of effects of reforms at the LSE in 1997 when the exchange removed the obligation of dealers to quote firm two-way prices for a subset of stocks, and allowed the public to compete directly with dealers in these stocks through the submission of limit orders on a centralised order book. They find that the limited pre-trade quote transparency significantly increased trading costs for large trades greater in size than the usable depth on the book, allowing dealers to charge higher effective spreads, and make bigger gross revenues for themselves.

Ángeles de Frutos & Manzano (4/2002) examine theoretically the impact of quote visibility by analysing two dealer market structures which differ in their degree of quote dissemination. Their key finding is that risk-averse dealers use less aggressive pricing strategies in the more transparent market, and this results in worse prices. Quote transparency alleviates uncertainty about the prices quoted by other dealers and, hence, reduces dealers’ need to compete for order flow.

Rindi (28/5/2002) studies theoretically the consequences of pre-trade transparency on liquidity in an order driven market. For a given proportion of informed agents, and the more transparent the market is, the more liquid it is. With endogenous information acquisition, however, this result can be reversed. Greater transparency reduces the incentives to acquire information, and thus the equilibrium number of informed agents who enter the market. This in turn can reduce liquidity, because informed agents can accommodate the liquidity shocks of liquidity traders, since they are not exposed to adverse selection. However, when agents are competitive, transparency increases liquidity as it reduces uninformed traders’ adverse selection costs. The effects of transparency on liquidity are therefore ambiguous. Transparency does however, increase informational efficiency and volatility.

4.3. Trades

A relatively detailed summary of the literature examining the effects of post-trade transparency on market performance and best execution is presented here. As previously, the literature is described in chronological order. The articles span a range of different market structures, jurisdictions and contexts, and also employ all three of the analytical techniques of theory, experimentation, and empirical analysis.

Roëll (10/1988) studies theoretically the effect of allowing dealers who are not required to publish details of their trades, to compete with market-makers who are required to do so. The less regulated dealers are always able to under-cut the more regulated market-makers initially, and are
then able to use the confidential knowledge gained from dealing with the informed trader, to withdraw from the loss-making side of the market at a later stage. This option is not, however, open to the more regulated market-makers.

Kyle & Roëll (1989) investigate theoretically the effects of implementing post-trade transparency in a dealer market theoretical model. When there is no transparency, dealers compete to invest in information by dealing with an informed trader, and therefore their spreads are relatively tight initially. Once a dealer has traded, however, he has an informational advantage over both his competitors and the liquidity traders, and they will lose out in subsequent trading. Post-trade transparency has two opposing effects: bid-ask spreads decrease because information about transactions reaches all market participants; but they widen because market makers have less incentive to compete to capture the information which a trade with the informed trader will bring. When averaged over all transactions, liquidity traders benefit from post-trade transparency, while informed traders lose.

Nasdaq undertook a study of a range of NASDAQ OTC stocks which were designated as National Market System (NMS) securities on 8/2/1983, a status which brought with it full post-trade transparency. Nasdaq found that the volume of trading increased significantly for most of the stocks that were upgraded to NMS status. It could not conclude, however, that this result arose unequivocally as a result of the changes in reporting rules. No significant changes in bid-ask spreads were observed for the upgraded stocks before and after their change of status, and similarly there appeared to be no apparent pattern in the changes of volatility. A similar analysis was undertaken by Seguin (9/1992). He found that daily returns volatilities declined by 8-10% immediately upon the inception of transaction reporting, bid-ask spreads were roughly 3.5% lower in the post NMS period, and required rates of return also declined. These effects were not, however, homogenous across all stocks.

Madhavan (1995) compares theoretically a “consolidated” market with full trade publication and a “fragmented” market, with no trade publication. He finds that if dealers are given a choice as to whether to disclose their trading information or not, they will not do so, even if there are potential economies of scale in consolidating order flow. This is because such fragmentation reduces price competition. Fragmentation also benefits large liquidity traders and informed traders by yielding lower execution costs. The effects of non-disclosure on noise traders is ambiguous. If trade disclosure is mandatory for at least one dealer, but voluntary for the rest, the unconstrained dealers will again choose not to disclose trading information. The expected trading volume of such dealers is larger than for the constrained dealers, as the constrained dealers are not able to match the quotes of the non-disclosing dealers. The relaxation of transparency rules can also improve the market’s performance, for example by lowering spreads. The fragmentation which arises as a result of a lack
of transparency does lead, however, to higher price volatility and less price efficiency.

Lyons (1996) examines theoretically what level of post-trade transparency a group of competing dealers will choose to establish in a market similar to the foreign exchange market. He identifies two opposing effects of transparency. On the one hand it accelerates the revelation of true prices, while on the other it impedes dealers’ management of their inventory risk. At one extreme, too noisy a public signal about inter-dealer order flow provides customers too little incentive to trade initially, and if they do not trade, they cannot share in the risk that would otherwise be borne entirely by dealers. At the other extreme, full transparency ensures that the interim price reveals all order flow information. In this case all price risk has been absorbed by the dealers already, so subsequent customer risk-sharing is irrelevant. Dealers prefer an environment with incomplete transparency. A moderately noisy signal slows price adjustment and allows time for customers to trade, thereby sharing the dealers’ risk.

Flood, Huisman, Koedijk, Mahieu & Roëll (3/1997) study experimentally how post-trade transparency affects the performance of a multiple-dealer market. A key result is that post-trade transparency significantly improves the informational efficiency of price quotes. This appears to be because the enhanced transparency reduces profit opportunities for informed traders, as their information is more rapidly transmitted to uninformed market participants. Uninformed traders, be they dealers or customers, thus gain at the expense of informed traders. On average, traders also face significantly smaller effective dealing spreads in the transparent market, as good prices are easier to find. In contrast, opening spreads are wider in the transparent market than in the opaque one. This difference disappears over time, however, because as trading proceeds more information is progressively revealed, and the information differences between opaque and transparent markets disappear.

Naik, Neuberger & Viswanathan (7/1999) examine theoretically whether the disclosure of public trade details improves the welfare of a risk-averse investor in a dealership market, where a market-maker executes a public trade and then off-sets its position by trading with other market-makers via an inter-dealer market. They distinguish between quantity risk and price revision risk. Quantity risk arises because the investor faces an endowment shock and has to readjust his asset holdings. Price revision risk arises when the investor negotiates with the dealer and in the process reveals private information about the true value of the asset. Greater transparency reduces adverse selection and thus leads to improved sharing of quantity risk. Greater transparency, however, also obliges the public investor to bear more of the price revision risk. The welfare benefits of greater transparency are thus ambiguous.

Bloomfield and O’Hara (1999) show experimentally that post-trade transparency enhances the informational efficiency of transaction prices. However, it also causes dealers’ opening spreads
to widen dramatically, apparently by reducing market makers’ incentives to compete for order flow. Subsequently, however, trade disclosure has little effect on spreads, which rapidly decline to zero. As a result, trade disclosure benefits market makers at the expense of both informed traders and liquidity traders who cannot time their trades. Such transparency, however, has no welfare effect on active traders who can wait until spreads narrow sufficiently to undertake their trades. Quote disclosure has little effect on either allocational efficiency, informational efficiency, or bid-ask spreads.

The International Stock Exchange (1989) summarises a study the exchange undertook concerning some stocks for which enhanced post-trade transparency had been implemented. No significant increase in turnover for the promoted shares was found. MacIntyre (1991) examines the effects of the LSE’s 1991 rule changes governing transparency on quote quality, turnover, and depth. For the most liquid stocks, the rule changes lead to an increased amount of information about trading volume being published, but a decreased amount about transaction prices being disseminated. For the medium-sized stocks, price and size transparency was increased, while for the smaller stocks it remained the same. The effects of the new transparency regime were mixed. For the more liquid stocks, individual dealer spreads and market bid-ask spreads remained relatively constant, while for the stocks for which transparency increased, spreads and touches declined. For the larger stocks there was a decline in total trading volume, while for the smaller stocks there was an increase. For those stocks for which transparency improved, there was a significant decrease in the number of large trades undertaken.

Gemmill (12/1996) explores the way in which share prices on the LSE reacted to large trades under three different trade publication regimes. Gemmill’s main conclusion was that delayed publication had surprisingly little impact on spreads, speed of adjustment, smoothing or the ultimate price level. He did note, however, that although there was an advantage to knowing that a large trade occurred, the information of the existence of a transaction leaked out so rapidly after its occurrence that the potential advantage could not be exploited.

Board & Sutcliffe (1/1995a) undertook an empirical investigation of the effects of transparency on the LSE. Their main conclusion concerning transparency was that market makers did not consistently use the opportunity offered by the transparency regime to offset the excess inventory acquired as a result of a large trade. In addition, they found evidence of a small but permanent price impact after large trades, information of which was not published immediately to the market.

Following the implementation of an enhanced transparency regime for the LSE in 1/1996, Board & Sutcliffe (1996) undertook a further study to analyse the effects of these rule changes. They find that market-makers did not make full use of the publication delay to unwind their
positions, and that the change in the publication rules resulted in a large reduction in the value of trades subject to delayed publication, without any accompanying reduction in the number of block transactions or any widening of bid-ask spreads. Saporta, Trebeschi & Vila (18/5/1999) similarly find no effect of the increase in post-trade transparency on the LSE in 1/1996 on either the distribution of trades or spreads, or on the adverse selection and inventory-holding cost components of the traded spread.

Porter & Weaver (1998) study the effects of post-trade transparency reporting on Nasdaq. They find that many more trades are reported out-of-sequence relative to centralised exchanges such as the NYSE and AMEX, and suggest that this late-trade reporting is neither random nor the result of factors (such as “fast” markets, lost tickets, and computer problems) outside Nasdaq’s control. Porter & Weaver note that the trades most likely to be reported late are large block trades, especially those at away prices, suggesting that late-trade reporting is beneficial to Nasdaq dealers.

4.4. Linkages

A range of studies have attempted to assess the manner in which competition between two or more markets trading the same or associated assets is affected by transparency.

Mulherin, Netter, & Overdahl (1991a) & (1991b) analyse whether exchanges’ restrictions on the activities that the purchasers of their prices and quotes are allowed to undertake are anti-competitive by looking at some historical evidence. They summarize the interpretation of this evidence as follows: “(1) Technological innovation is often accompanied by stricter definition of property rights. (2) Financial innovation arises from stronger definition of property rights. (3) There is an interaction between monitoring the behaviour of exchange members and limiting off-exchange trading of members.” They then argue that “technology, by itself, will not lead to lower transaction costs. Instead, new technology, combined with well-defined and well-enforced property rights, can lead to cost-reducing innovations”. Far from being anti-competitive practices, therefore, they conclude that exchanges’ restrictions on who could have access to their price and quote information, and on what terms this could be done, were the very mechanisms that lead to increased levels of trading and the use of advanced technology.

Pagano & Roëll (7/1991) find that the London Stock Exchange had derivative pricing with respect to the Paris and Milan bourses in those shares which were traded both in London and respectively Paris or Milan: prices in London tended to follow the prices obtaining on the domestic continental markets. Prices were, however, very tightly arbitraged despite the lack of any post-trade transparency on SEAQ International.

Kofman & Moser (3/1996) investigate the nature of the competition between two futures exchanges, the DTB and LIFFE, which both traded the same contract but which had different levels
of transparency - the DTB was a more anonymous and less transparent automated system, while LIFFE was a less anonymous and more transparent pit-based exchange. They find that the realised spreads on both systems are virtually identical, but that the components of the spreads on both exchanges are quite different. An order processing costs component is found to be significantly lower on the DTB than on LIFFE, while the costs that arise as a result of information asymmetries and time-varying expected returns are significantly higher on the DTB. The difference in transparency between the two systems appears to have no effect on the lead/lag relationship in pricing. Both markets appear informationally efficient.

Drudi & Massa (1/2002) study how dealers behave when there exist parallel markets for the same asset that are characterised by different degrees of transparency. They show that the optimal trading strategy may involve price manipulation. Specifically, informed dealers may refrain from trading in the more transparent market in order to exploit their informational advantage in the less transparent one, or they may use the more transparent market in order to manipulate prices. Both the strategies, however, increase market depth in the more transparent market. Drudi & Massa confirm the empirical implications of their model in the Italian Treasury bond market.
Appendix 1. Notes

1 Amongst the literature summaries which are not referred to in Appendix 2, are Board, Sutcliffe & Wells (9/2002), Ganley, Holland, Saporta, & Vila (Spring 1998), Holland (5/2000), and Sabatini & Tarola (5/2002).

2 See Domowitz (1993).

3 Stoll (8/2000) studies the sources of trading friction, and identifies and characterises seven distinct empirical measures.

4 Sometimes informed traders are called “insiders”, however this terminology is confusing as it has no direct link with the legal concept of “insider trading” which in most securities markets is illegal.

5 As has been recognised in the academic literature and also by a range of different regulatory authorities. See, for example, Bacidore, Ross & Sofianos (1999), Blume (2002), EAMA (6/2002), Foucault (6/2002), FSA (4/2001), Harris (9/3/1996), Levitt (4/11/1999), and Macey & O’Hara (7/1997).

6 Lee (Ch. 4, 1998) discusses a range of different ways in which the term fragmentation may be applied to securities markets.


9 Amongst the data that may be disseminated are: the price of last trade; the quantity of last trade; time of last trade; the identities of parties to last trade; the high, low, opening, & closing, trade prices; aggregate price data and price indices; cumulative trade volume; best bid and ask quotes; quantities at best bid and ask quotes; identities of parties placing these quotes; bid and ask quotes behind the best quotes; quantities at these quotes; identities of parties placing these quotes; high, low, opening, & closing, mid-quote prices; requests for quotes; identities of parties who requested quotes; number of individuals logged onto system; and identities of those individuals. See Domowitz (1993) and IOSCO (6/1990).

10 Proponents of a Consolidated Limit Order Book who argue strongly against fragmentation, include, Amihud & Mendelson (1991), Mendelson & Peake (1979), and Mendelson, Peake & Williams (1979). See also Porter & Thatcher (Spring 1998).

11 For further discussions about fragmentation, see Ch. 4, Lee (1998), and SEC (14/7/1992) & (23/2/2000) on the topic, and some of the comments the SEC received, including Schwartz, (5/4/2000), Scott Miller (26/5/2000), Weaver (27/4/2000), and Wood (18/4/2000).

12 Subsequent studies sought to better the methodologies employed in this earlier work, by controlling for firm-specific and trading system attributes [for example, Affleck-Graves, Hegde & Miller (1994) & Christie & Huang (6/1994)].

13 See, for example, Schwartz & Steil (1996) and (2000).

14 Bacidore, Battailo & Jennings (2001) do not, however, find evidence of this occurring on the NYSE.

15 See also Booth, Lin, Martikainen & Tse (7/2002).

16 The author was not able to read Hamet (8/2002) or Hansch, Naik & Viswanathan (10/1999). Hamet examines the effects of internalisation by dealers on the LSE of trading in French
shares. She suggests that LSE market makers seem to divert a new clientele to the Paris Bourse, which increases both market activity and the breadth of the Bourse’s order book. Hansch, Naik & Viswanathan discuss preferencing and internalisation in the UK context.

17 Kandel & Marx (2/1999) present a model of Nasdaq which examines the possibility that broker-dealers may practice internalization or preferencing. They do not, however, investigate the effects of these different market structures on market performance or best execution.


19 See also Booth, Lin, Martikainen & Tse (7/2002).


21 Rindi (28/5/2002) reports that Albanesi & Rindi (2000) show that the introduction of anonymity increased liquidity on the MTS.

22 The International Stock Exchange was what the LSE was called at that time. Few details of the study have been provided.
Appendix 2. Excerpts from Other Summaries

A range of summaries of the literature on the issues of fragmentation, internalisation and transparency have been prepared, other than the one presented in this report. Key excerpts from some of these summaries are presented here so as to obtain some different viewpoints about what the main lessons of the literature are.

Competition, and Fragmentation

Harris (forthcoming 2002, Ch. 26) summarises the main issues about consolidation and fragmentation as follows:

Markets consolidate because traders attract traders. Trading is easiest and cheapest where most traders of an instrument or similar instruments trade. Liquidity attracts liquidity.

Markets fragment because the trading problems that traders solve, differ. Different market structures serve some traders better than others. Markets fragment when, for enough traders, benefits from differentiation exceed benefits from consolidation.

Some traders are small and unconcerned about the price impacts of their trades, while other traders are large and very concerned about front-running. Small traders prefer market structures that widely expose their orders so that everyone can see and react to them. Large traders prefer market structures that allow them to control how and to whom their orders are exposed.

Some traders are well informed about fundamental values and therefore very concerned about revealing their information, while others are relatively uninformed and very concerned about minimising transaction costs. Uninformed traders prefer markets where they can be identified and given better prices. Informed traders prefer consolidated markets with anonymous trading so that they can hide in the order flow.

Some traders are impatient to trade and therefore willing to pay for liquidity, while others are patient and willing to wait for their price. The former prefer quote-driven markets, while the latter prefer order-driven markets.

Notwithstanding these differences, all traders appreciate the benefits of consolidation. Traders often trade in markets that they do not like simply because those markets are most liquid. Conversely, no market will attract and keep liquidity if it does not provide good service to many traders. Competition among market structures generally reveals the market structures that best serve various types of traders.

Fragmented markets consolidate when traders can access information about market conditions within each segment. Traders use this information to adjust their orders, reroute their orders, or issue new orders. Prices and liquidity in each segment thereby reflect information from all other segments.

Traders naturally enforce price priority in segmented markets when they seek the best prices for their orders. Traders do not enforce secondary order precedence rules, such as time precedence, across market segments. Only
coordinated regulation can implement such rules.

Fragmented markets generally will provide less regulatory oversight than is socially optimal. Good regulatory activities benefit everyone, but exchanges can charge only those traders who trade in their segments. Only coordinated regulation can ensure that markets provide adequate regulatory oversight.

Stoll (6/5/2002) summarises the debate about consolidation and fragmentation as follows:

The forces of centralisation are two-fold - one on the supply side and one on the demand side. First, on the supply side, a market reaps economies of scale in processing transactions. The average cost of trading a share of stock declines with the number of shares traded. As a result, the first mover into the trading business has a great advantage because it can process trades at lower cost than a competitor using the same technology. Second, on the demand side, a market generates network externalities. A market is a communications network, and like other networks, ... its attractiveness depends on the number of others on the network. Traders want to trade where other traders are already trading because the probability of a successful trade is a function of the number of other traders using the market. Consequently, network externalities, like economies of scale, lead to a first mover advantage.

Several factors have made competition from satellite markets more effective in recent years and have weakened the centralising forces of economies of scale and network externalities. First, the transparency of quotes and transaction prices makes it possible for a satellite market to credibly guarantee that the price in the primary market is being matched. For many years, the NYSE jealously guarded its price information and limited the dissemination of its quotes and transaction prices. Without knowledge of where the price is, investors prefer the primary market where price discovery takes place. With transparency, a trader can be assured the price in a satellite market at least matches the price in the primary market.

Second, satellite markets not only match prices, but they also pay for order flow from brokers. A typical payment might be one or two cents per share for market orders from retail investors that are judged to be uninformed. Payment is not made for limit orders or for order flow judged to be informed. Payment for order flow has been criticised because the payment goes to the broker, not to the customer whose order is being routed to the satellite market. While payment for order flow is quite common among satellite exchange[s], it is not necessarily sufficient to overcome the natural centralising forces. If the primary market is the low cost producer of transaction services, it can make the same payment.

Third, technological change has made competition more effective. Nimble new exchanges may be able to implement new, low cost, electronic trading systems more quickly than existing markets and thereby attract order flow away from established markets. Communications technology also reduces the switching costs of moving trades from one market center to another. The ease with which orders can be routed to a satellite market has improved.

Fourth, regulatory policy in the U.S. has fostered competition and fragmentation. The SEC has required greater transparency, which enhances competition from new markets. Second the SEC has required markets to
link, which has given satellite markets access to the primary market. Such links enable dealers in the satellite market to lay off inventory in the primary market and provide an opportunity for brokers to route orders to the satellite market.

It is not evident how the conflict between centralisation and fragmentation will be resolved in the future. The forces of centralisation - economies of scale and network externalities - are strong. While they have been weakened by technology and regulation, they have not been weakened to the extent that markets will necessarily fragment into many separate unconnected market centers. If markets do fragment, the adverse consequences are small because markets are linked by high speed communications systems. The term “fragmentation” has a harmful connotation, but, in fact, fragmentation is just another word for competition. Competition among markets is a good thing because it fosters innovation and efficiency. Separate markets may exist, but when linked by high-speed communications systems they act almost as one.

The cost of fragmentation is that priority rules are difficult to maintain across markets. Price priority can usually be maintained because, with transparency, the investor can send his order to the market with the best price. But even price priority can sometimes be violated, for example, when large orders in one market trade through prices in another market. Time priority is likely to be violated as traders prefer to trade in one market over another that may offer the same price.

**Internalisation**

Harris (forthcoming 2002, Ch. 25) summarises the main effects of internalisation as follows:

Brokers internalise and preference order flows to extract value from largely uninformed orders that execute at wide spreads. Payments for order flow ensure that dealers in perfectly competitive wholesale dealing markets do not obtain excess profits from trading these orders at wide spreads. The commissions and other order flow inducements that brokers must offer their clients to obtain their orders ensure that brokers in perfectly competitive retail brokering markets do not excessively profit from internalisation or payments for order flow. When competition is perfect in retail and wholesale order flow markets, low commissions offset poor execution so that net prices do not ultimately depend on best execution standards.

Competition in no market is perfect, however. Dealers and brokers with market power will exploit that power and ultimately obtain excess profits from public traders. How much excess profit they obtain depends on how competitive these markets are.

Wholesale dealers have some market power by virtue of the economies of scale associated with their operations. These economies have led to substantial consolidations through mergers and acquisitions in the dealing segment of the trading industry. Although the remaining firms undoubtedly increased their market power, the economies of scale have also made them more efficient. These economies of scale make it difficult for new entrants to compete aggressively in the wholesale order flow market.
Generally, the more convoluted is a competitive system, the less efficient it will be. The wholesale and retail order flow market system is a more complex competitive system than is a centralised market. We therefore can presume that it will be less efficient.

By taking orders away from common market mechanisms, internalisation, preferencing, and internal order crossing practices make it harder for natural buyers and sellers to find each other. Internalisation and preferencing also weaken central markets by reducing incentives to quote aggressively. These practices therefore must ultimately increase the total transaction costs of all buy side traders. Internalisation and preferencing, however, probably provide small uninformed traders with better net prices - spread plus commission - than they would otherwise obtain. Internal order crossing likewise provides many traders with services that exchanges and brokers would not otherwise provide…

Macey & O’Hara (1997) summarise a key issue about best execution in the context of internalisation as follows:

When markets compete in different ways with respect to the different components of trade execution, it is no longer so clear what “best” execution is, let alone how or when it is attained. Moreover, numerous market practices such as preferencing and payment for order flow, which presumably arise as competitive responses to established markets, directly challenge the provision of a “best” execution standard. Well-meaning attempts to mandate best execution as a consumer-protection device run counter to attempts to make markets less centralised and more competitive. We argue that this difficulty makes best execution both un-wieldy and unworkable as a mandated legal duty: pursuing a narrow concept of best execution may make markets less competitive.

The SEC (15/4/1997) summarises the main effects of preferencing and internalisation as follows:

After undertaking a thorough analysis of the practice of “preferencing,” … the Commission has concluded that preferencing has not had a deleterious effect on the national market system. To the contrary, market data for preferencing exchanges is at least as good as that of other regional exchanges, and in some cases, is better. Moreover, preferencing has furthered the ability of the CSE and BSE to compete in the national market system. Furthermore, the Commission has concluded that preferencing is not necessarily inconsistent with the best execution of customer orders.

Through a review of the trading practices permitted under the rules of the various exchanges, the Commission has found that preferencing is but one method to accommodate the internalisation of order flow. … internalisation exists in all markets and has manifested itself in many variations over the years. Even now, both primary exchange and regional exchange rules permit a specialist to engage in a significant amount of dealer activity and provide a number of means through which time priority among orders can be avoided. Further, preferencing is merely one means utilised by broker-dealer firms to capture market maker profits from their customer order flow.

Internalisation is made possible by the lack of time priority for orders across
exchanges. The absence of such intermarket time priority allows broker-dealers to route their customer order flow to a particular exchange and, in many instances, execute such orderflow as principal regardless of “same-priced orders or quotations entered prior in time” on other exchanges. Preferencing programs merely allow this to occur on an intra-exchange basis. CSE’s preferencing program and BSE’s CSI are the latest manifestations of internalisation. As these preferencing programs have made both exchanges more competitive, they have attracted criticism from competitors.

In recent years, the Commission has examined practices such as preferencing and other inducements to order flow and found that such practices are not necessarily inconsistent with best execution of customer orders. Nevertheless, a broker-dealer that automatically routes order flow to a particular exchange has the responsibility to regularly and rigorously evaluate the executions received by customers in that market. Such an evaluation takes on particular importance when preferencing or other inducements for order flow are involved.

The Commission undertook an analysis of CSE and BSE firms’ order handling practices and compared those to other arrangements between broker-dealers and either their affiliated or joint venture specialist units. The Commission found no significant variance among the order handling procedures of preferencing and non-preferencing firms. Broker-dealer firms frequently route their retail order flow to captive market makers. However, firms generally review execution quality across various market centers on a regular basis. As noted above, such a review is a necessary prerequisite to ensure that preferencing is not interfering with the firm’s best execution obligation.

… The Commission could find no evidence from the measures of market quality it analysed to suggest that the CSE’s preferencing program and the BSE’s CSI have had a harmful impact on investors and the national market system. The Commission’s analysis found that the CSE’s quotation spreads were narrower than those of the other regional exchanges. The BSE’s quotation spreads for stocks were wider than the CSE’s but generally comparable to those of the other regionals. The CSE’s quotation depth was found to be better than that of the other regionals, except when the CSE was at both sides of the NBBO. In such instances, it was comparable to that of most regionals. The BSE’s quotation depth was equivalent to, or exceeded that of, other regionals when the BSE was at one or both sides of the NBBO. However, the BSE was only at one or both sides of the NBBO 5% of the time. The Commission believes that the BSE’s quote performance should improve as a result of its implementation of quote entry capability for competing specialists in March 1997, and its future implementation of automatic routing of orders to the competing specialist at the ITS/BBO with priority. The Commission will continue to monitor the BSE’s progress in this area. In light of the foregoing, the Commission concludes that the adoption of preferencing programs generally has not harmed the market quality of both the CSE and BSE, while making these exchanges more competitive.

… The data indicates that preferencing has not necessarily diminished the ability of customers to acquire quality executions of their orders. … Both
the CSE and BSE guarantee the execution of orders up to a specified size at the ITS/BBO or better. The size of this execution guarantee is significantly higher than the average size of an individual investor’s order. Importantly, the rate of price improvement (i.e., the percentage of time an order is executed at a better price than the best quotes prevailing in the market at the time the order arrived at the market or market center) on the CSE surpasses that of the other regionals. In addition, the CSE’s rate of price improvement compares favorably to that of the NYSE. The BSE’s rate of price improvement in preferred stocks is as good as or better than that of most other regionals…

… The Commission’s analysis found that the limit order fill rates (i.e., the percentage of time limit orders on an exchange receive an execution) on the CSE and BSE are higher than those of the NYSE. Further, the CSE’s fill rate is usually better than that of other regionals. Moreover, the Commission found that limit orders sent to the CSE and to the BSE in CSI stocks usually were executed as quickly as those sent to the NYSE and more quickly than those sent to the other regionals. The Commission also found that, prior to the adoption of the Commission’s Display Rule, the CSE and BSE had in place policies with regard to the exposure of customer limit orders to trading interest on their respective exchanges and to the national market system. Moreover, the Commission believes that the Display Rule will further enhance the exposure of customer limit order interest not only on the CSE and BSE, but on all exchanges, thereby leading to greater order interaction within the NMS. This increased order interaction will result in quicker and more frequent executions of customer limit orders in all markets.

… The Commission found no evidence of adverse costs for preferred orders in the aggregate vis-a-vis other markets in the national market system. Further, the Commission’s analysis has found no evidence that customers necessarily receive inferior executions of their orders on the CSE. Indeed, in almost all measures of market quality examined, the CSE and BSE performed as well as other regional exchanges and, in some cases, better than other markets. Accordingly, our analysis did not find that preferencing has harmed the quality of executions on the CSE or BSE. [The data collected by the Commission was used for the preferencing study, and should not be relied upon by broker-dealers in making order routing decisions. The Commission’s data is already several months old and predates the effective date of the Order Execution Rules. In addition, the Commission analysed selected measures of market quality in the aggregate. Broker-dealers, in regularly and rigorously evaluating execution quality across markets, should not use this report as a substitute for their own monitoring, but instead should conduct their own, current analyses of execution quality based on the types of orders they handle.]

Although the Commission does not believe that the preferencing of customer orders necessarily results in inferior executions, the Commission notes that such programs do raise significant agency-principal concerns. As the Commission stressed in the orders permanently approving the CSE’s preferencing program and the BSE’s CSI, the Commission believes that exchanges operating preferencing programs and broker-dealers participating in such programs need to continuously monitor the executions received by customers whose orders are preferred. Specifically, broker-dealers must regularly and rigorously evaluate the executions received by customers
whose orders are preferred. While to date the Commission has found that these programs in aggregate do not appear to have had an adverse effect on either the executions received by customers whose orders are preferred or the national market system as a whole, these findings should not be taken to mean that the Commission believes that such adverse effects may not arise in the future. Indeed, changed circumstances may result in a determination by the Commission that the findings it has made in this study require reconsideration. For example, a significant increase in the amount of preferencing activity as a percentage of overall national market system activity that results in a decline in execution quality on the national market system, or a significant deterioration in the surveillance and regulatory programs of an exchange operating, or broker-dealers participating in, a preferencing program would require reconsideration. Therefore, the Commission will continue to review the effects of the practice of preferencing, as well as those of other exchange practices that allow for time priority to be avoided, on the national market system on an ongoing basis to ensure that this practice is consistent with the maintenance of fair and orderly markets, the protection of investors and the public interest, and the furthering of the national market system goals of Section 11A of the Act.

Transparency

Biais, Chester & Spatt (8/3/2002) summarise the main effects of transparency as follows:

the first generation of the market microstructure literature, analys[ed] the price impact of trades and the spread, assuming competitive suppliers of liquidity. Under this assumption, the revenues of the agents supplying liquidity, corresponding to the spread, simply reflect the costs they incur: order-handling costs (Roll, 1984), adverse-selection costs (Kyle, 1985, Glosten and Milgrom, 1985, Glosten, 1994) and inventory costs (Stoll, 1978). While this literature identified these costs theoretically, it also developed empirical methodologies to analyse data on transaction prices and quantities and estimate trading costs, through the relation between trades and prices and the bid-ask spread (Roll, 1984, Glosten and Harris, 1988, and Hasbrouck, 1988). This literature has shown that trades have both a transitory and a permanent impact on prices. While the former can be traced back to order-handling and inventory costs, the latter reflects information, although not necessarily private information. Furthermore, as data on inventories became available, empirical studies of specialists’ or traders’ inventories examined the relevance of the inventory paradigm. While this literature has shown that inventory considerations have an impact on the trades of liquidity suppliers, the empirical significance of the impact of inventories on the positioning of their quotes is less clear.

While the theoretical literature emphasises the benefits of transparency to mitigate frictions due to asymmetry and strategic behaviour, empirical and experimental results are much more ambiguous.

In transparent markets abundant information is available to investors and traders about orders and quotes (ex-ante transparency) and about transactions (ex-post transparency). As this tends to equalise information
across market participants, transparency reduces the magnitude of adverse selection problems. Since these problems reduce the gains from trade, transparency can be anticipated to increase welfare. Indeed, within the context of an adverse selection-based model of the spread, Pagano and Roell (1996) show theoretically that transparency reduces the transaction costs incurred by uninformed investors. Consistent with that analysis, Flood, Huisman, Koedjick and Mahieu (1999) find that pre-trade transparency narrows spreads in experimental financial markets.

One could argue, however, that trade disclosure can make it harder to supply liquidity to large traders. After large trades, in a transparent markets, the market maker can be in a difficult bargaining position to unwind his inventory. Naik, Neuberger, and Viswanathan (1999) offer an interesting counterargument. After the risk-averse dealer has bought a block from a potentially informed trader, he seeks to unload his position. Yet to mitigate his price impact, he reduces the size of his trade, thus reducing his ability to share risk. This does not arise with trade disclosure. In that case, since the market has already taken into account the information content of the trade, the dealer can unwind his inventory with little incremental price impact. Consequently, trade disclosure enhances risk sharing. The empirical evidence in Gemmill (1996) is consistent with the view that transparency at least does not reduce liquidity. Gemmill (1996) analyses liquidity in the London Stock Exchange under three publication regimes: from 1987 to 1988 dealers had to immediately report their trades, from 1991 to 1992 they had to do so within 90 minutes, while from 1989 to 1990 they had 24 hours to do so. He finds that there is no gain in liquidity from delayed publication of block trades, as the spreads and the speed of price adjustment are not affected by the disclosure regime.

Yet, in a dynamic trading environment, transparency can have ambiguous consequences, as shown by the experimental and theoretical analyses of Bloomfield and O’Hara (1999 and 2000). Consider a two-period model where informed agents can submit market orders to one of N liquidity suppliers. If the market is opaque, only the liquidity supplier who accommodated the order at the first period is informed about its occurrence, sign and size. Consequently, at the first period, liquidity suppliers quote relatively tight spreads, to attract order flow, and thus acquire private information. At the second period, however, the liquidity suppliers who did not participate in the initial trade face a double winner’s curse problem: i) with respect to the informed agent, and ii) with respect to the informed liquidity supplier. This relatively severe adverse selection problem widens their spreads. The market spread is wide also, as the informed liquidity supplier finds it optimal to undercut his competitors by just one tick. Thus, different temporal patterns emerge in the transparent and the opaque market. While in the latter, spreads may be initially relatively large, they decrease fast, as information is revealed through time. In the former, in contrast, while initial spreads are relatively tight, later spreads tend to remain relatively high.

O’Hara (undated) summarise some key effects of transparency as follows:

“economic experiments provide disquieting evidence that transparent markets may be less liquid than markets with weaker reporting requirements. Transparency reduces the information content of specific trades and so
reduces dealers’ incentive to compete for orders. As a result, bid-ask spreads in transparent markets tend to be wider than those in less transparent markets. This accords with the experience in actual markets. Spreads on Instinet, for example, are frequently narrower than those on NASDAQ. The “Golconda exchange” [defined as an exchange that provides optimal liquidity] may be less transparent than some of the markets that currently dominate global trading.

Madhavan (2000) summarises the main effects of transparency as follows:

First, there is broad agreement that transparency does matter; it affects the informativeness of the order flow and hence the process of price discovery. Greater transparency is generally associated with more informative prices. Second, complete transparency is not always beneficial to the operation of the market. Indeed, many studies demonstrate that too much transparency can actually reduce liquidity because traders are unwilling to reveal their intentions to trade. Third, there is also general agreement that some disclosure, as opposed to no disclosure whatsoever, can improve liquidity and reduce trading costs. Finally, changes in transparency are likely to benefit one group of traders at the expense of others. The literature almost uniformly agrees that traders who trade on private information signals will prefer anonymous trading systems while liquidity traders, especially those who can credibly claim their trades are not information-motivated, prefer greater disclosure. Consequently, no single market structure is viewed as best by all parties.

Gravelle (undated) summarises the main effects of transparency as follows:

It seems that there exists no optimal market structure, in terms of transparency, that would benefit all types of market participants, nor accommodate, at the same time, increased market efficiency and liquidity. As such, the current state of research calls into question the view that increased transparency is unambiguously beneficial for markets, and suggests that market structure and investor attributes should be important considerations in assessing the appropriate transparency regime.

Stoll (6/5/2002) summarises the main effects of transparency as follows:

Transparency refers to the disclosure of quotes (at which trades can take place) and of transaction prices (at which trades did take place). The NYSE displays only the top of the book, that is the best bid and ask, but not the other orders on the book. The ECNs display the entire book. The benefits of transparency are three-fold. First, transparency speeds price discovery and enhances market efficiency, for with transparent markets all investors see the current quotes and the transaction prices, and no investor trades at the wrong price. Second, transparency helps customers monitor brokers. The public dissemination of quotes and transactions allows a customer to determine that his transaction is in line with others at the same time. Third, transparency enhances competition, for it allows competing dealers to guarantee the best price anywhere, but do it at a lower commission or lower spread. The costs of transparency arise from adverse incentive effects. First, traders may be reluctant to place limit orders, particularly if they are large, because the display may convey information that will make the price move against the limit order. Second, display of limit orders may make it easier for traders to exercise the free trading option and thus reduce the incentive
to place limit order. If no one knows whether a limit order exists, it is more
difficult to pick it off, but if the limit order is displayed, it can be more
readily picked off.
Appendix 3. References


Battalio, Robert H., Robert Jennings, and James P. Selway III. The Relationship between Market-making Revenue, Payment for Order Flow, and Trading Costs for Market Orders. Georgia State University, University of Cincinnati, Kelley School of Business Indiana University, Goldman Sachs (2/2000).


Flood, Mark D., Ronald Huisman, Kees C. G. Koedijk, Ronald Mahieu & Ailsa Roëll. *Post-Trade Transparency in Multiple Dealer Financial Markets*. Concordia University - Faculty of Commerce & Administration, Erasmus University, Erasmus University Rotterdam (EUR) - Department of Finance, Erasmus University and Princeton University - Department of Economics (3/1997).


