Workup

Romans Pancs*

University of Rochester

Abstract

In pure limit-order markets, the use of large orders is discouraged by potential front-runners. This problem can be mitigated by using expandable orders or iceberg orders, or by splitting a large order into smaller ones. An expandable order gives a trader an option to sequentially expand the size of his trade while holding the price fixed. An iceberg order enables a trader to commit to a maximal trade size at some price, without fully revealing that size to other traders. This paper provides a theoretical model of expandable orders and the accompanying quantity negotiation, called (also by practitioners) the "workup." The workup is ubiquitous in the U.S. and Canadian bond markets and in over-the-counter markets. The model suggests that even when iceberg orders are available, traders will still use the workup if splitting an order is prohibitively costly, front-running is a threat, and the exchange lacks trust.

1 Introduction

In financial markets, a **limit order** is a trader's public commitment to buy or sell a specified quantity at a specified price or better. One problem with the pure limit-order markets is that the use of large limit orders is discouraged by potential front-runners. A **front-runner** is a trader who, having observed a posted limit order to sell a large quantity, anticipates the price impact of the impending sale, sells a certain amount ahead of that sale, and then, after the price impact of the

*Department of Economics, University of Rochester, Rochester, NY 14627 (email: rpancs@gmail.com). The feedback from the associate editor and the two anonymous referees has greatly improved the paper.

impending sale has been realized, buys back his sold amount, at depressed prices, thereby realizing a profit. One can mitigate the front-running problem by using **iceberg orders** (which are as limit orders except that some of the quantity remains invisible to other traders until the initially visible quantity has been executed) or **expandable orders** (which are as limit orders except that the initially specified quantity can be increased further after the buyer and the seller have been matched).[1] This paper provides a theoretical model of expandable orders and the accompanying quantity negotiation, called the **workup**.

The model addresses the following questions:

1. What determines traders' strategies during the workup?

2. When can expandable orders be replaced by iceberg orders?

These questions are not addressed in the existing literature, which is mostly empirical. Empirically, workups have been studied by Huang et al. (2002), Boni and Leach (2004), and Dungey et al. (2009). Theoretically, iceberg orders have been studied by Anand and Weaver (2004), Esser and Mönch (2007), Aitken et al. (2001), Buti and Rindi (2009), and Moinas (2010), but these orders' normative relationship to the workup has not been explored.

The posed questions are important because the workup is ubiquitous. It is used in the U.S. and Canadian government bond markets and in over-the-counter markets.[2] Most trades in the U.S. government bond market occur in exchanges called electronic communication networks (ECNs), such as eSpeed and BrokerTec.[3] These ECNs have evolved in response to traders' demand for anonymity and for greater privacy of their intended trades.[4] The orders are expanded often; workups leading to larger than initially specified quantities account for two thirds of the dollar volume of trade in these ECNs (Dungey et al., 2009). In over-the-counter markets, workups are often used by brokers, who assist in negotiations between the counterparties whose identities are not revealed to each other.

[1] Alternatively, if neither iceberg nor expandable orders are available, a trader can dynamically submit a sequence of small limit orders, as in the model of Kyle (1985), instead of submitting a single large limit order.

[2] D'Souza and Gaa (2004) discuss the Canadian bond market. Duffie (2011) analyzes over-the-counter markets and observes that workup is used in broker-mediated over-the-counter markets.

[3] The institutional details of BrokerTec and eSpeed are described by Mizrach and Neely (2006), Harris (2003), Kuznetsov (2006), and Schwartz et al., eds (2006).

[4] Mizrach and Neely (2006) estimate that the treasury market trading volume in the third quarter of 2005 was $473 billion a day, of which 70 percent were traded in ECNs.

The theoretical properties of the workup are studied in a bilateral-trade model, introduced in Section 2. In the model, a seller and a buyer exchange an asset for money. The per-unit transaction price is fixed. The negotiation is only about the quantity. During the workup, traders alternately increase the desired quantity by a given amount until either trader vetoes a further increase. Then the agreed-upon amount is traded.[5] Asymmetric information is of two kinds. The seller's private information, or type, is his desired trade size, called **block**. The buyer's type is whether he can front-run; if he can, he is called a **front-runner**; otherwise, he is called a **dealer**.[6]

By assumption, the dealer wishes to buy as much as he can. The front-runner does not; he does not value the asset. If the front-runner accepts quantity increases during the workup, he does so only to learn something about the seller's desired trade. This information helps the front-runner decide whether he will profit if he front-runs. Front-running can be profitable because of the assumption that if the seller fails to sell his desired quantity within the workup, he will sell the remaining quantity outside the workup, thereby causing a price impact. The front-runner exploits this price impact by selling high (before the price impact) and buying low (after the price impact). Front-running can be unprofitable due to the fixed cost it entails. The seller prefers not being front-run, because selling outside the workup entails both a price impact and a fixed cost.

Traders' strategies in the workup are studied in Section 3, which solves for an equilibrium of the workup game. The focus is on a particular equilibrium (described in Theorem 1), which minimizes the front-runner's expected payoff (as shown in Theorem 2). In this equilibrium, called the **gradual-disclosure equilibrium**, the negotiated quantity increases at each step of the workup by the same, minimal feasible increment. These minimal increments ensure that as many seller's types as possible take the same action for as long as possible, thereby impairing the front-runner's inference about the seller's block. This impaired inference reduces the front-runner's expected payoff, which could be justified as a desideratum in a richer model with participation costs, in which the payoff reduction might discourage the front-runner from joining the exchange, and so from incurring the socially wasteful cost of front-running and from forcing the seller to incur the

[5] An equivalent interpretation of the workup is that each increase in quantity is traded immediately after it has been agreed upon.

[6] Burdett and O'Hara (1987) were the first to study the information effects arising from desired trade sizes when the trades are not motivated by private information about the asset's common value. Vayanos (2001) observes that if large trades were motivated by private information about the asset's common value, large traders would outperform small (poorly-informed) ones—a prediction not supported by data. BrokerTec and eSpeed platforms have been specifically designed for dealers, who are large traders.

socially wasteful cost of selling outside the workup.

The seller's preference for offering the smallest quantity increment at every round relies on specific equilibrium beliefs. This preference is driven by the front-runner's belief that any deviation from the equilibrium strategy comes from the seller with the largest possible block, whom it is profitable to front-run. One particular deviation is the seller's offer of his entire block at the first round. This deviation is equivalent to the submission of a limit sell order. The seller's unwillingness to disclose his block at the first round is thus equivalent to unwillingness to submit a limit order for his entire block, in the presence of expandable orders.[7,8]

At the analyzed gradual-disclosure equilibrium, the front-runner also optimally offers the smallest possible increment at each round. He does so in order to mimic (and thus be mistaken for) the dealer and thereby keep acquiring information about the seller's block.

Theorem 3 derives the model's testable implications by studying the dependence of the front-runner's equilibrium strategy on the model's parameters. For instance, as long as the front-runner participates in the workup, he stays in it for longer if front-running is costlier, if the amount he can sell when front-running is less, or if the price impact from sale is smaller. In all these cases, front-running is less profitable, and so the front-runner needs more information before deciding whether to front-run. He acquires this additional information by staying in the workup for longer to ascertain that the seller's block is sufficiently large.

Section 3 concludes by addressing a potential problem with the workup. The workup's gradual-disclosure equilibrium (and indeed, any equilibrium) relies on the exchange's ability to coordinate traders on specific off-equilibrium-path beliefs. Theorem 4 introduces an alternative mechanism, dubbed the **button mechanism**, that constrains traders to propose the smallest possible increments at every round. As a result, the workup's gradual-disclosure equilibrium outcome is the button mechanism's only equilibrium outcome; the exchange is no longer required to be able to coordinate traders.

Section 4 explores the relationship between the workup and iceberg orders. The goal is to understand when the workup can be supplanted by iceberg orders. The approach is to use mech-

[7] The model's implication that expandable orders help avoid the price impact is supported by Fleming and Mizrach (2009), who find that the price impact in BrokerTec is small. Boni and Leach (2004) report similar results for voice protocols.

[8] Madigan and Stehm (1994) report interviews with traders in the U.S. government bond market. These traders confirm that the protection of the private information about the intended trades is a reason to seek anonymity.

anism design to find the best trading mechanism, to show that this mechanism can be implemented using iceberg orders, and to observe that the assumption of the exchange's commitment not to abuse the confidential information contained in the submitted iceberg orders is crucial for traders' willingness to submit these orders. The workup does not rely on such commitment, and therefore may prevail in some environments.

The results in Sections 3.4 and 4 can also be interpreted normatively. In particular, if the exchange's goal is indeed to minimize front-runners' payoffs, the button mechanism is superior to the regular workup. By contrast, if the exchange's goal is to maximize the sum of the traders' payoffs, the exchange may wish to retain the regular workup and leave it to the traders to coordinate on the appropriate equilibrium. Furthermore, Section 4 suggests that if the exchange can commit not to abuse the private information contained in the traders' submitted orders, then the workup can be improved upon by a static trading mechanism (Theorems 5 and 6). For the exchange to design this mechanism optimally and for traders to behave in it optimally, however, all must agree on the economic environment's fine details. If these details are unavailable, the exchange may prefer to opt for the workup.

Related Literature Gradual disclosure in the workup is related to games of timing. When the workup has a gradual-disclosure equilibrium, the reduced-form workup game is a "simple timing game," as discussed by Fudenberg and Tirole (1991, Section 4.5). In the workup game, the agreed-upon quantity is the counterpart of time in a timing game; a trader chooses a threshold quantity after which he rejects further increases. The payoff structure resembles a preemption game (Harris and Vickers, 1985), a special case of the simple timing game.

Because each trader has private information that can affect the other trader's payoff, the workup resembles bargaining with correlated values, first studied by Evans (1989) and Vincent (1989). In the workup, however, traders negotiate about quantity, not price. Moreover, the interdependence of values emerges indirectly, through traders' actions and outside options, not directly, through their valuations for the asset. Finally, both traders are privately informed.

Some features of the equilibrium play in the workup resemble bargaining with a commitment type. Myerson (1997, Section 8.8, Theorem 8.4) analyzes a bargaining model in which the commitment type accepts any offer that is at least as good as a certain threshold offer. The commitment type is mimicked by the "regular" type, who thereby secures a payoff that is close to what the commitment type demands. (Abreu and Gul, 2000, derive a related result in a model with multiple commitment types.) In the workup, the dealer can be interpreted as a commitment type, whom the front-runner (the regular type) wishes to mimic in order to continue acquiring information about the seller's block.

A workup effectively "sells" to the front-runner information about the seller's trading intention. The front-runner's "payment" is his loss from liquidating the asset, which he does not value. Admati and Pfleiderer (1990) study the sale of information in a rational-expectations equilibrium, whereas Hörner and Skrzypacz (2009) do so in a game-theoretic framework. The economic environment, the designer's objectives, and the class of admissible mechanisms differentiate the present paper from Hörner and Skrzypacz (2009). Both papers find, however, that, in equilibrium, private information is released gradually.

The workup is a mechanism that caters to traders with multi-unit demands. Compared to auctions with multi-unit demands (pioneered by Wilson, 1979), the workup abstracts from price determination and has a special payoff structure motivated by the study of front-running. Kremer and Nyborg (2004) study complete-information common-value uniform-price auctions for divisible quantities. In their auctions, each bidder can submit only finitely many bids, each of which must specify a marginal quantity proportional to a given quantity multiple and a price proportional to a given tick. The seller's revenue increases as the tick size decreases, because competing by undercutting each other becomes cheaper for bidders. By contrast, altering the minimal quantity increment in the workup would have no competitive ramifications, because each party to the workup faces no competition on the same side of the market. Instead, reducing the minimal quantity increment reduces the amount of information that the front-runner can glean at each round of the workup, thereby tending to reduce the front-runner's equilibrium payoff.

The model of the workup assumes minimal quantity increments but does not explain them. By contrast, Kastl (2012) proposes a model in which the discreteness of bids arises endogenously. He studies private-value uniform-price and discriminatory auctions for divisible goods, and shows that if specifying additional price-quantity pairs is costly for a bidder, the bidder restricts his bids to only a few pairs, and that the loss from this restriction is rather small.

The critical feature of the workup model is the threat of front-running. Front-running is ubiquitous in practice and is of concern to both individual traders and regulators, as emphasized by

Harris (1997). Front-running often emerges in models of dual trading, in which a broker-dealer observes his client's order before trading on both his own account and on his client's account—that is, in a dual capacity (Grossman, 1989; Röell, 1990; Fishman and Longstaff, 1992; Pagano and Röell, 1993; Danthine and Moresi, 1998; Bernhardt and Taub, 2008). Front-running may also emerge when a trader observes an announcement of a so-called "sunshine" trader, who broadcasts his trading intention before submitting a request to trade (Admati and Pfleiderer, 1991). In either case, a trader who observes another's trading intention and can trade first (as in the models of Fishman and Longstaff, 1992; Pagano and Röell, 1993; Brunnermeier and Pedersen, 2005; Bernhardt and Taub, 2008) may profitably front-run—in the same manner as the front-runner does in the present paper. By contrast, if all trade simultaneously, front-running is infeasible, and the knowledge of another's trading intention may even be profitless (Rochet and Vila, 1994).

The closest empirical study is Boni and Leach (2004). They analyze the GovPX platform, which features expandable orders, and corroborate the rationale for the use of expandable orders that is proposed by the workup model. In their own words, they "document evidence consistent with the hypothesis that traders use expandable limit order strategies to reduce costs associated with stale limit orders and information leakage"—for instance, many trades exceed the initially quoted quantities—with the caveat that "expandable limit orders do not completely eliminate pre-trade information leakage or the potential for information free-riding." Consistent with this evidence, the workup's gradual-disclosure equilibrium reduces the information leakage about the seller's block by initially concealing its size, but does not eliminate this leakage, which enables the front-runner to profit from the information learned during the workup. At odds with the evidence, however, the gradual-disclosure equilibrium has the largest feasible number of rounds, whereas Boni and Leach (2004) document just two rounds on average. This discrepancy may be specific to the GovPX's historic data, which are based on mediation by human brokers, who are presumably slower than the algorithms in present-day ECNs.

2 Model

A buyer and a seller negotiate the exchange of an asset for money. If the seller fails to sell his intended quantity during the negotiation (called "workup"), he sells the remainder with a delay and experiences a price impact, represented by a strictly decreasing and smooth inverse demand function $p : \mathbb{R} \to \mathbb{R}$. The buyer faces the same inverse demand function if he front-runs.

Traders: The seller (**S**) must sell a privately known quantity, or **block**, z for exogenous reasons, at any price in the domain of interest. Let $z \in Z \equiv \{\delta, 2\delta, .., \bar{z} \mid \delta > 0\}$, where $\bar{z}$ is a multiple of some minimal increment δ with $\delta < \bar{z} < \infty$.[9] The buyer (**B**) privately knows whether he is a front-runner (**F**), with probability $\pi \in (0,1)$, or a dealer (**D**), with probability $1 - \pi$. **B** assigns a positive probability $f(z)$ to z. The cumulative distribution function corresponding to f is denoted by F. A trader's **type** is his private information.

Actions: The agreed-upon quantity by the end of round $n \geq 1$ of the workup is denoted by Q_n. During the **workup**:

1. Set $n = 1$ and $Q_0 = 0$.

2. Each trader alternately is either a proposer or a responder. At round $n = 1$, **S** is the proposer.

   2.1 At round n, the proposer either **rejects** further increases or **offers** $q_n \geq 0$.

   2.2 The responder **accepts** or **rejects**. The proposer accepts or rejects.

   2.3 If someone rejects, quantity Q_{n-1} is traded at price $p(0)$. Otherwise, increment $Q_n = Q_{n-1} + q_n$ and $n = n + 1$, and go to step 2.1.

After the workup:

3. **F** liquidates his position (if any) and either front-runs or does not front-run. (The exact meaning of liquidation and front-running will become clear once their effect on payoffs has been described.)

4. **S** sells $z - Q_{n-1}$ at prices given by the demand curve $p(\cdot)$.

The traders observe each other's actions. **S** cannot offer or accept any quantity that is infeasible given his block z.

Strategies: Each trader's strategy specifies the planned rejection round and the sequence of workup offers, both functions of his type and of the actions' history. **F**'s strategy, in addition, specifies $a \in \{0,1\}$, where $a = 1$ if **F** front-runs, and $a = 0$ if **F** liquidates without front-running. **D** cannot front-run.

Payoffs: Each trader is an expected-utility maximizer.

[9] Increment δ may be a technological restriction, which, if irrelevant, can be dispensed with by letting $\delta \to 0$.

D's payoff is strictly increasing in the amount bought from **S** and is additive in money, as will be motivated.

At cost $c > 0$, **F** can use up to amount $k > 0$ of the asset for front-running. Let $Q \geq 0$ denote the equilibrium amount traded in the workup. **F**'s realized payoff is:[10]

$$-p(0)\,Q + \int_0^Q p(s)\,\mathrm{d}s + a\left(\int_0^k [p(Q+s) - p(z+s)]\,\mathrm{d}s - c\right).$$

The payoff's first term is **F**'s expenditure on buying the amount agreed upon during the workup. This amount is then sold, or **liquidated**, at prices given by $p(\cdot)$, hence the revenue captured by the second term. The third term is the profit from **front-running**, which equals the revenue from short-selling k before **S** sells $z - Q$ less the expenditure on buying back k at the prices depressed by **S**'s sale and less the cost of front-running, c. (Because **F**'s payoff is increasing in k, when front-running, **F** optimally uses up the entire amount k available to him.)

To study the non-degenerate case, assume

$$\int_0^k [p(s) - p(\bar{z}+s)]\,\mathrm{d}s > c, \tag{1}$$

implying that **F** would front-run if he knew that **S** would sell amount $\bar{z}$.

If $Q < z$, **S** incurs an order-resubmission cost $r > 0$ and sells $z - Q$ at prices given by $p(\cdot)$. Thus, **S**'s payoff is:

$$p(0)\,Q + \mathbf{1}_{\{z>Q\}}\left(\int_0^{z-Q} p(Q + ak + s)\,\mathrm{d}s - r\right), \tag{2}$$

where $\mathbf{1}_{\{\}}$ is the indicator function. The payoff's first term is **S**'s revenue from selling Q in the workup. The second term is the revenue from selling $z - Q$ at prices possibly depressed by **F**'s front-running.[11]

Game: The traders, the set of their possible strategies, and the traders' payoffs induce the workup game.

Discussion of assumptions: The independence of **S**'s intended trade from the price is a simplifying assumption, which can be justified in applications. For example, **S** may be a pension

[10] When $Q = 0$, **F** does not trade with **S** and either front-runs immediately or does not front-run at all.

[11] When $Q = 0$, **S** sells z right away, bypassing the workup.

fund selling a predetermined number of bonds to disburse pensions. Alternatively, **S** can be an insurance company selling bonds to compensate its clients for damages.

D's payoff, increasing in the amount of his trade, is motivated by his unmodelled gain from intermediation over a long horizon. In a richer model, one can imagine **D** engage in many workups over time, with different counter-parties, some of whom sell and some of whom buy, all at the long-run price $p(0)$. **D** would earn zero from traders, whereas his preference for larger trades can be induced by rebates, often paid by exchanges to encourage liquidity provision.[12] In this richer model, **F**'s trade at a transient price impact would be motivated by his impatience and reluctance to carry inventory (which, incidentally, are empirical characteristics of high-frequency traders).

The inverse demand function $p(\cdot)$ can be generated by many small uninformed liquidity traders who have high demand for immediacy (i.e., must trade immediately). The assumption that $p(\cdot)$ is strictly decreasing can be motivated by (unmodelled) limited mobility of capital, which may be due to search frictions,[13] decreasing marginal utility from holding the asset, or inventory effects.

F's cost of front-running, c, is motivated by the cost of borrowing k units of the asset and by the opportunity cost of time devoted to front-running. In the model, this cost generates an option value for delaying front-running.

S's cost r is motivated by the cost of order formulation and submission, delayed execution, and privacy lost through trading with multiple partners. In the model, this cost ensures that, during the workup, **S** increases the agreed-upon traded quantity up to z in a bet that **B** is **D**, not **F**. Hence, throughout the paper, r is assumed to be sufficiently large.[14]

Traders have no asymmetric information about the asset's common value. This assumption is appropriate for the Treasuries market, populated by large sophisticated traders.

[12]For instance, in 2009, the New York Stock Exchange introduced rebates to so-called supplementary liquidity providers.

[13]Duffie and Strulovici (2011), in a paper studying the effects of capital mobility on asset pricing, cite existing theoretical and empirical literature on capital-market imperfections in reinsurance, corporate bond, and equity markets. The authors conclude, "The typical pattern suggests that the initial price response is larger than would occur with perfect capital mobility, and reflects the demand curve of the limited pool of investors that are quickly available to absorb the shock."

[14]The intuition for the problem in which r is small and **S** can split his orders over time (which the present model does not allow) is familiar from Kyle (1985). The present paper is concerned with the complementary problem.

3 An Equilibrium with Gradual Disclosure

This section is concerned with the positive inquiry into the circumstances in which expandable orders are used, and with the determinants of traders' strategies during the workup. Theorem 1 illustrates when **S** prefers to disclose his block only gradually. Theorem 2 suggests why such gradual disclosure may appeal to the exchange. Theorem 3 describes how **F**'s equilibrium strategy depends on the model's parameters.

3.1 The Equilibrium Disclosure Theorem

It will be shown that the workup game has a perfect Bayes-Nash equilibrium in which **S** discloses his block fully but gradually, in minimal increments. Given **S**'s block z, an on-equilibrium-path offer profile $[q_n(z)]_{n\geq 1}$ is **gradual** if $q_n(z) = \delta$ for each n with $Q_{n-1} < z$—that is, each trader offers the smallest feasible quantity increment δ until **S** exhausts his block. A perfect Bayes-Nash equilibrium with a gradual offer profile is a **gradual-disclosure equilibrium**. Its existence is established in the following theorem. In the theorem's statement, the amount of the asset **revealed** by **S** at round n is the agreed-upon amount Q_{n-1} plus the round-n offer made or accepted by **S**.

Theorem 1. *For a sufficiently small π or a sufficiently large r, the workup game has a gradual-disclosure equilibrium. On the equilibrium path, the play is as follows. At each round, **D** offers and accepts δ until **S** rejects. **S** offers and accepts δ either until **F** rejects or until **S**'s entire block has been agreed upon, whereupon **S** rejects. **F** either (i) rejects **S**'s offer outright, at round 1, and does not front-run, or (ii) at each round, offers and accepts δ until the amount revealed by **S** reaches some $\hat{Q} \geq \delta$ (whereupon **F** rejects, liquidates, and front-runs) or until **S** rejects (whereupon **F** liquidates and does not front-run).[15] Moreover, if either **F** follows type-(i) strategy and $n = 1$ or **F** follows type-(ii) strategy and the agreed-upon amount Q_{n-1} is less than $\hat{Q} - \delta$, then **S** (strictly) prefers offering δ to offering any other amount.*

Proof. See Appendix. □

Theorem 1 admits as a special case the outcome in which **F** rejects the first trade δ that **S** offers, and does not front-run. This case is captured by type-(i) strategy and prevails when **F**'s cost of front-running is high or when he is sufficiently confident that **S**'s block is small. Another

[15] Section 3.3 analyzes the properties of $\hat{Q}$.

special case is when **F** rejects the first trade in the workup and front-runs immediately. This case is captured by type-(ii) strategy with $\hat{Q} = \delta$, and prevails when **F**'s cost of front-running is small or when he is sufficiently confident that **S**'s block is large.

When **F** adopts type-(ii) strategy with $\hat{Q} > \delta$, **F** agrees to some trades, even though doing so entails a loss due to a price impact at liquidation. **F** agrees to trade with **S** because **S**'s gradual acceptance of larger trades reveals information about z. In particular, having agreed upon some amount $Q - \delta$ and having observed **S**'s offer to trade δ more, **F** learns that $z \geq Q$. If **F**'s prior belief is that **S**'s block is distributed with a "fat tail," the knowledge that $z \geq Q$ might convince **F** that z is sufficiently high to justify front-running. **F**'s loss from trading is the cost he chooses to pay for information.

When **F** adopts type-(ii) strategy with $\hat{Q} > \delta$, **S** with a block $z \geq \hat{Q}$ gradually reveals his entire block, instead of revealing $\hat{Q} - \delta$ and thus precluding front-running, because he cannot rule out **B** being **D**, who never front-runs. Rejecting further increases after having revealed less than $\hat{Q}$ will avert front-running when **B** is **F**, but it will also force **S** to sell the remainder of his block at a price impact and at a fixed cost r. When π is sufficiently small or r is sufficiently large, behaving as if **B** is **D** is optimal for **S**. That is, gradually revealing the entire block is optimal.

Even though **S** reveals his block eventually, he would not do so immediately, unless front-running were inevitable (i.e., unless **F** adopted type-(ii) strategy with $\hat{Q} = \delta$). **S** never offers anything but δ at equilibrium. Hence, if **S** were to offer, say, his entire block $z > \delta$ immediately, at round 1, then **F** would believe (it is assumed) that **S**'s block is sufficiently large to justify immediate front-running. Gradual disclosure enables **S** to avoid this outcome.

S's offering his entire block z at round 1 is equivalent to submitting a limit order. Theorem 1 thus exhibits an equilibrium at which **S** always weakly, and sometimes strictly, prefers submitting an expandable order to submitting a limit order.

As the preceding discussion of disclosure at round 1 suggests, the on-equilibrium-path play is sustained by off-equilibrium-path beliefs that are not uniquely determined by the model. Essentially, the theorem's proof assumes that **S** believes that a deviating **B** is **F**, and **B** believes that a deviating **S** has $z = \bar{z}$. These beliefs discourage deviations. In particular, **S** would reject **B**'s any offer other than δ, believing that this deviant offer is made by **F**, who will front-run as soon as **S** accepts. Analogously, **S** refrains from offering anything other than δ, because a deviant offer would make **F** believe that **S** has the largest possible block, thereby inducing **F** to front-run. The concern about the indeterminacy of off-equilibrium-path beliefs and the multiplicity of equilibrium outcomes is central to Section 3.4.

Expandable orders can help **S** prevent front-running even if **F** does not engage into the workup. One can construct examples in which, at the gradual-disclosure equilibrium, **F** does not engage into the workup and does not front-run. If expandable orders were unavailable and **S** had no choice but to reveal his block by submitting a limit order, **F** would sometimes front-run.

The equilibrium in Theorem 1 may be focal because its on-equilibrium-path play is simple. **S** simply adheres to gradual disclosure, without having to know **B**'s belief f. **F**, in turn, either never front-runs or adheres to a front-running strategy that is described by a single real number, $\hat{Q}$.

3.2 The Gradual-Disclosure Equilibrium is Front-Runner Pessimal

In some circumstances, the gradual-disclosure equilibrium is the exchange's most preferred one. In particular, assume that the exchange minimizes **F**'s expected profit, for reasons to be explained shortly. A trivial way to achieve this outcome is not to conduct any trades at all, in which case **F**'s payoff will be zero. To rule out this trivial outcome, assume that the exchange restricts attention to **admissible equilibria**, in which **D** buys **S**'s entire block.[16] A **front-runner pessimal equilibrium** minimizes **F**'s expected payoff over all admissible equilibria of the workup game.

Theorem 2. *If a gradual-disclosure equilibrium exists, it is front-runner pessimal.*

Proof. See Appendix. □

The exchange's objective to minimize **F**'s payoff can be motivated by the desire to discourage **F** from participating in the exchange. This motivation emerges if the present model is augmented to acknowledge that (i) **F**'s probability of participation is increasing in his expected payoff (e.g., because there is a fixed cost of participation), and (ii) if **F** does not participate, he is replaced by **D** (who may already be in the market but is sometimes preempted by **F**).[17] In such a model,

[16] The requirement that **D** buy **S**'s entire block can be motivated by **D**'s unmodelled participation cost, which requires that **D**'s payoff, and hence trading volume, be large enough.

[17] Feature (ii) can be justified on empirical grounds. In practice, high-frequency traders (the counterparts of **F** in the model) buy and resell within milliseconds, preempting regular traders (**D**, in the model). The welfare gains from a trade executed a millisecond sooner are dubitable, whereas costs from displacing **D** (c and r in the model) may be substantial. (For one of the early reports on high-frequency trading, see "Stock Traders Find Speed Pays, in Milliseconds," The New York *Times*, 24 July 2009. See also "Markets: Ghosts in the machine," *Financial Times*, 17 February 2010.)

discouraging **F** from participation conserves the socially wasteful front-running cost c when **F** would have front-run and conserves the socially wasteful cost r of resubmitting an order when **S** would not have sold his entire block to **F**. Even the minimized payoff can be large relative to **F**'s outside option, however, in which case **F** may still be willing to participate, and the exchange may prefer to pursue a different objective, such as surplus maximization, for instance.

In the proof of Theorem 2, a front-runner pessimal equilibrium is found first by analyzing a relaxed problem in which **S**'s and **D**'s incentives are neglected and the focus is on **F**'s incentives. The relaxed problem is **F**'s decision problem in which an offer profile is revealed to him round by round, and he must decide, at each round, whether to accept or liquidate, and then whether to front-run. The proof proceeds by showing that **F**'s payoff is minimized by the offer profile that has no gaps and no jumps, which are defined as follows.

A profile has a **gap** if, at some round, for some block z, **S** accepts zero or offers zero. A profile has a **jump** if, at some round, for some block z, **S** offers or accepts an amount greater than the minimal increment δ. An equilibrium with an on-equilibrium-path offer profile that has no gaps and no jumps is a gradual-disclosure equilibrium.

Intuitively, the gradual-disclosure equilibrium is front-runner pessimal because it minimizes the information that **F** learns about z, for any agreed-upon amount of trade. By removing jumps and gaps, the exchange pools the maximal number of **S**'s types who are willing to trade any given quantity, thus minimizing the informativeness of **S**'s acceptance of any given increment. (The removal of a jump adds rounds at which **F** can act, but this addition does not improve **F**'s payoff, as the theorem's proof shows.) In **F**'s decision problem, poorer information translates into a weakly lower payoff.

In conclusion, note that if the exchange could not offer expandable orders (i.e., if the workup were unavailable) and had to resort exclusively to limit orders, the front-runner's expected payoff would weakly (and sometimes strictly) rise. Indeed, his payoff would be maximized. To see this, suppose that a limit order were the only kind of order that **S** could submit, and that **B** could either accept that order (by submitting a buy limit order) or reject it (by submitting no order). Then, for a sufficiently large r or a sufficiently small π, **S** would optimally submit the limit order to sell his entire block. **D** would accept. **F** would reject and would front-run if and only if **S**'s order size exceeded a certain threshold. **F**'s payoff would be maximized because without trading with **S**, he would learn z and take a privately optimal front-running decision.

3.3 Comparative Statics

This section collects some of the model's testable implications under the assumption that the equilibrium described in Theorem 1 prevails, also as parameters vary. It is analytically convenient to consider the case of an arbitrarily small δ, corresponding to the case in which the asset is divisible.[18] The changes in the parameters with respect to which comparative statics exercises are performed can affect equilibrium existence. Therefore, it is assumed that these changes are sufficiently small, so that the gradual-disclosure equilibrium continues to exist. The focus is on **F**'s strategy because it is the one that varies as parameters vary.

When **F** adopts type-(ii) strategy (as defined in Theorem 1), the equilibrium threshold $\hat{Q}$ is found by analyzing **F**'s payoff function, which solves a difference equation. Let x denote the amount revealed by **S**. Let $V\left(x,\hat{Q}\right)$ denote **F**'s continuation payoff from following type-(ii) strategy. This payoff is defined recursively:

$$V\left(x,\hat{Q}\right)=\frac{1-F\left(x\right)}{1-F\left(x-\delta\right)}V\left(x+\delta,\hat{Q}\right)-\frac{f\left(x\right)}{1-F\left(x-\delta\right)}\int_{0}^{x}\left(p\left(0\right)-p\left(s\right)\right)\mathrm{d}s,\qquad x<\hat{Q},\tag{3}$$

with the terminal condition:

$$V\left(\hat{Q},\hat{Q}\right)=-p\left(0\right)\left(\hat{Q}-\delta\right)+\int_{0}^{\hat{Q}-\delta+k}p\left(s\right)\mathrm{d}s-\mathbb{E}_{z|z\geq\hat{Q}}\left[\int_{0}^{k}p\left(z+s\right)\mathrm{d}s\right]-c.\tag{4}$$

Equation (3) expresses **F**'s expected payoff as the difference between the expected continuation payoff if **S** reveals his willingness to trade δ extra units of the asset, less the expected cost of buying x from **S** and then liquidating. Equation (4) is the expected payoff from front-running once the revealed amount x reaches the front-running threshold $\hat{Q}$.

The optimal threshold $\hat{Q}$ solves $\max_{s\in Z}V\left(\delta,s\right)$. Because Z is finite, the solution exists. If $V\left(\delta,\hat{Q}\right)\geq 0$, **F**'s optimal strategy is indeed type-(ii). If $V\left(\delta,\hat{Q}\right)<0$, **F** optimal strategy is type-(i).

An arbitrarily small minimal increment is modelled by setting $\delta\to 0$. The support of the

[18] Technically, the focus on the environment with $\delta\to 0$ is valid because the existence of the gradual-disclosure equilibrium for some δ implies the existence of a gradual-disclosure equilibrium for any smaller δ (that divides $\bar{z}$ without a remainder).

distribution of **S**'s blocks becomes an interval, $[0, \bar{z}]$, and the difference equation (3) becomes a differential equation. With some abuse of notation, assume that the limit of **B**'s belief is described by a positive p.d.f. f, with the associated c.d.f. F. Then, at the beginning of the workup, **F**'s expected payoff induced by type-(ii) strategy with threshold $\hat{Q}$ is $J\left(\hat{Q}\right) \equiv \lim_{\delta \to 0} V\left(\delta, \hat{Q}\right)$, or

$$J\left(\hat{Q}\right) = R\left(\hat{Q}\right)\left(1 - F\left(\hat{Q}\right)\right) - \int_0^{\hat{Q}} \int_0^t \left[p\left(0\right) - p\left(s\right)\right] f\left(t\right) \mathrm{d}s\mathrm{d}t, \tag{5}$$

where

$$R\left(\hat{Q}\right) \equiv -p\left(0\right)\hat{Q} + \int_0^{\hat{Q}+k} p\left(s\right)\mathrm{d}s - \int_{\hat{Q}}^{\bar{z}} \int_0^k p\left(z+s\right) \frac{f\left(z\right)}{1 - F\left(\hat{Q}\right)} \mathrm{d}s\mathrm{d}z - c. \tag{6}$$

F's expected payoff $J\left(\hat{Q}\right)$ equals his payoff from front-running, denoted by $R\left(\hat{Q}\right)$ and discounted by the probability that **S** trades z before **F** front-runs, less the expected discounted loss due to buying **S**'s entire block and then liquidating it, without front-running.

An interior optimal threshold $\hat{Q}$ satisfies the first-order condition obtained by equating to zero the derivative of (5) with respect to $\hat{Q}$:

$$c\frac{f\left(\hat{Q}\right)}{1 - F\left(\hat{Q}\right)} - \left[p\left(0\right) - p\left(\hat{Q}+k\right)\right] = 0, \tag{7}$$

where the first term is **F**'s value of information learned by agreeing to trade an extra unit of the asset; if **F** agrees to an extra unit, then, with probability $f\left(\hat{Q}\right) / \left(1 - F\left(\hat{Q}\right)\right)$, **S** rejects, in which case **F** saves c by not front-running. The second, bracketed, term is the cost of trading an extra unit, which is the difference between what **F** pays for the unit, $p\left(0\right)$, and what **F** sells this unit for, $p\left(\hat{Q}+k\right)$. This bracketed term can be interpreted as the cost of information acquisition.

To assess how the threshold $\hat{Q}$ varies with the magnitude of the price impact, index the demand function by a parameter b. Let the indexed demand function $p_b\left(s\right)$ be strictly submodular in b and in the size of trade s.[19] In other words, the price impact is increasing in b.

The following theorem collects comparative statics results. The theorem adopts the convention according to which $\hat{Q} = \infty$ means that **F**'s optimal strategy is type-(i), and $\hat{Q} < \infty$ means that **F**'s optimal strategy is type-(ii).

Theorem 3. *Let $\delta \to 0$ and let **F**'s equilibrium strategy be as in Theorem 1 for all parameter values*

[19]Strict submodularity means that $p_b\left(s\right) - p_b\left(s'\right)$ is increasing in b for $s' > s$.

considered below. Then, the optimal front-running threshold $\hat{Q}$ is

(i) weakly increasing in the front-running cost c;

(ii) weakly decreasing in the front-running amount k;

(iii) weakly decreasing in the magnitude b of the price impact, provided $\hat{Q} < \infty$.

Proof. The proof proceeds in two steps. Step 1 determines when $\hat{Q}$ is finite. Step 2 derives the dependence of $\hat{Q}$ on parameters when $\hat{Q}$ is finite.

Step 1 By inspection of (5) and (6), R, the terminal payoff from front-running, is strictly decreasing in c and strictly increasing in k, which also implies that $\max_s J(s)$ is strictly decreasing in c and strictly increasing in k. Thus, $\max_s J(s) > 0$ (i.e., immediate liquidation without front-running is suboptimal, or $\hat{Q} < \infty$) if and only if $c < \hat{c}$ for some $\hat{c}$ (holding all other parameters fixed) or if and only if $k > \hat{k}$ for some $\hat{k}$ (holding all other parameters fixed). In words, $\hat{Q}$ is finite as long as c is sufficiently small or k is sufficiently large, consistent with parts (i) and (ii) of the theorem.

Step 2 Consider the derivative

$$J'\left(\hat{Q}\right) \equiv \frac{\partial J\left(\hat{Q}\right)}{\partial \hat{Q}} = cf\left(\hat{Q}\right) - \left(1 - F\left(\hat{Q}\right)\right)\left(p_b(0) - p_b\left(\hat{Q} + k\right)\right).$$

This derivative is strictly increasing in c and strictly decreasing in k, by inspection. It is also strictly decreasing in b because $p_b(0) - p_b\left(\hat{Q} + k\right)$ is strictly increasing in b (by strict submodularity). Hence, the front-runner's maximand J has strictly increasing differences in $\left(\hat{Q}, c\right)$, in $\left(\hat{Q}, -k\right)$, and in $\left(\hat{Q}, -b\right)$. Then, when $\hat{Q} < \infty$, the comparative statics conclusions reported in the theorem's parts (i), (ii), and (iii) follow by the Monotone Selection Theorem of Milgrom and Shannon (1994).

Combining Steps 1 and 2 establishes the theorem's conclusion. □

Intuitively, as c increases, front-running becomes less profitable relative to immediate liquidation at round 1. If immediate liquidation is nonetheless suboptimal, higher c encourages **F** to accept increases in the traded quantity for longer in order to acquire additional information before performing costly front-running. An analogous argument applies to the decrease in k.

A greater price impact (i.e., a higher b in p_b) on the one hand increases **F**'s gain from front-running, but on the other hand makes **F**'s trade with **S** costlier because the amount bought from **S** must be sold at lower prices. Hence, the effect of the price impact on the optimality of immediate liquidation is ambiguous. Conditionally on the suboptimality of immediate liquidation, however, a greater price impact unambiguously induces earlier front-running.

Theorem 3 suggests how to test the model. Suppose that observing front-running in the data is possible by inspecting trades that immediately follow a workup. Then, the model predicts that shorter workups are followed by more sizable instances of front-running, and that shorter workups are observed in the markets in which the price impact is greater.

Whether and when **F** front-runs depends on his belief about z. Say that **F uses an expandable order**, or **expands**, if he adopts type-(ii) strategy with $\hat{Q} > \delta$. One can construct examples in which **F** is more likely to expand when the probability distribution of z has a low expectation (so that immediate front-running is suboptimal), a fat right tail (so that learning can make front-running profitable), and a high hazard rate for low values of z (so that learning is fast). Moreover, **F** is more likely to expand when the inverse demand function is "more concave." These implications are testable. In the data, **F**'s trade can be distinguished from **D**'s trade because **F**'s use of an expandable order is immediately followed by **F**'s sale of the amount he has bought from **S**, whereas **D** retains his purchase.[20]

3.4 Full Implementation in the Button Mechanism

A problem with the workup is that, apart from the gradual-disclosure equilibrium, it may have other, undesirable for the exchange equilibria. Coordination on the (desirable) gradual-disclosure equilibrium is further complicated by the fact that it is sustained by off-equilibrium-path beliefs, which the exchange may have trouble inducing. To tackle these problems, this section proposes an alternative mechanism for expanding an order. Called the "button mechanism," it has the outcome of the workup's gradual-disclosure equilibrium as its unique equilibrium outcome. In the words of implementation theory, the button mechanism fully implements the gradual-disclosure equilibrium outcome of the workup.

The **button mechanism** has a "clock" that displays the agreed-upon quantity, beginning with 0 and incrementing it by δ at every round $n \in \{1, 2, ..., \bar{z}/\delta\}$. At round 1, **B** publicly accepts or rejects the first increment. At every round $n \geq 2$, **S** publicly accepts or rejects (e.g., by pressing a

[20] This identification strategy suggests that if the exchange wanted to discourage front-running, it could prohibit (or charge for) the trades that immediately followed the use of expandable orders.

"button") the latest increment, after which **B** publicly accepts or rejects the latest increment. All rejections are irrevocable and lead to the trade of the agreed-upon quantity at price $p(0)$. This trade is followed by **F**'s liquidation, the decision regarding whether to front-run, and **S**'s sale of the remainder of his block (if any remains). Each trader's strategy up to the point at which **F** decides whether to front-run is an integer designating a round at which to "press the button" to stop expanding the order—hence the name "button mechanism."

The button mechanism can be relied on to minimize the front-runner's expected payoff, without the need to coordinate the traders' off-equilibrium-path beliefs or even on-equilibrium-path actions:

Theorem 4. *For a sufficiently small π or a sufficiently large r, all perfect Bayes-Nash equilibria of the game induced by the button mechanism are of the following form. At each round, **D** accepts until **S** rejects. **S** accepts until his entire block has been agreed upon, whereupon **S** rejects. **F** either (i) rejects at round 1, and does not front-run, or (ii) at each round, accepts until the agreed-upon amount plus the latest proposed increment reaches some $\hat{Q} \geq \delta$ (whereupon **F** rejects, liquidates, and front-runs) or until **S** rejects (whereupon **F** liquidates and does not front-run). Moreover, the value of $\hat{Q}$ and hence the equilibrium outcome are as in a gradual-disclosure equilibrium of Theorem 1.*

Proof. It is a strictly dominant strategy for **D** to accept at every round.

Given **D**'s strategy, regardless of **F**'s strategy, it is optimal for **S** to accept until his entire block has been agreed upon. Indeed, if **S** accepts at round n, his expected payoff is at least

$$p(0)\,z - \pi\left(r + \int_0^{z-Q_{n-1}} \left(p(0) - p(k + Q_{n-1} + s)\right) \mathrm{d}s\right), \tag{8}$$

where Q_{n-1} is the agreed-upon amount by the end of round $n-1$. If, by contrast, **S** rejects, his expected payoff is at most the amount given in (10) and reproduced here for convenience:

$$p(0)\,z - \left(r + \int_0^{z-Q_{n-1}} \left(p(0) - (1-\pi)\,p(s) - \pi p\,(Q_{n-1} + s)\,\mathrm{d}s\right)\right).$$

S prefers acceptance to rejection if (8) exceeds (10), which occurs when r is sufficiently large or π is sufficiently small, mimicking the logic of the argument in step 1.(a).ii of the proof of Theorem 1.

Thus, any action that **S** can take in the button mechanism is consistent with an equilibrium path for some z. Hence, no off-equilibrium-path beliefs for **B** need to be specified. The justification of **F**'s strategy described in the theorem's statement mimics parts 3.(a).i and 3.(a).iii of the proof of the optimality of **F**'s strategy in Theorem 1 and is omitted. □

4 When Iceberg Orders Dominate the Workup

Iceberg orders enable a trader to partially conceal his trading intention—just as expandable orders do in the workup. (An **iceberg order** is a commitment to trade a certain amount of the asset at a certain price or better, with a caveat that other market participants initially see only a part of that total amount.) These orders are ubiquitous in practice, sometimes are observed alongside workups (e.g., on BrokerTec platform, according to Fleming and Mizrach, 2009), and sometimes are absent from exchanges that use workups. One may wonder under what conditions, if ever, mechanisms with iceberg orders dominate the workup. The understanding of such conditions, and of the consequences of their failure, helps understand when one should expect to see—and the exchange would gain from using—workups or iceberg orders, or both.

This section shows that a mechanism with iceberg orders dominates the workup if the exchange can commit not to abuse traders' private information—for example, not to front-run by using the information contained in the submitted orders. This conclusion is reached by solving for an exchange-optimal mechanism and showing that its outcome can be implemented using iceberg orders. Exchange optimality will be defined in two ways: as an outcome of the minimization of **F**'s expected payoff (which has been the focus so far) and as an outcome of the maximization of the traders' surplus. An optimal mechanism solves a mechanism design problem, which assumes the exchange's ability to commit. This ability is unnecessary for the workup to be feasible.

Attention is restricted to **direct admissible mechanisms**, in which:

1. Each trader confidentially reports his type to the exchange.
2. The exchange recommends to each trader whether to participate.
3. Each trader participates or leaves the mechanism—that is, the mechanism is **voluntary**.
4. The exchange enforces **S**'s sale of an amount Q at price $p(0)$ to **B**.

5. The exchange recommends to the participating **F** whether to front-run.

6. **F** liquidates and front-runs or does not front-run.

7. **S** sells the remainder of his block $z - Q$.

By the Revelation Principle, in the mechanism design problem, no loss of generality occurs in restricting attention to direct mechanisms in which traders report their types truthfully and abide by the exchange's recommendations.[21] Nor does a loss of generality occur in assuming that both traders participate, because the exchange can provide each trader with his outside option at no cost. The described direct mechanism assumes that the exchange can enforce trade between **S** and **B** but can only recommend to **F** how to trade with the traders modelled by the inverse demand curve p. That is, the exchange enforces the submitted orders, but cannot let these orders influence the execution of future orders—which is a plausible assumption, especially if these future orders are submitted to a different exchange. Finally, because traders report their types simultaneously and confidentially, no trader can surprise the other with an off-equilibrium-path action. Hence, the exchange need not worry about off-equilibrium-path beliefs supporting the intended on-equilibrium-path play.

4.1 A Front-Runner Pessimal Mechanism

A **front-runner pessimal mechanism** is a direct mechanism that minimizes the front-runner's expected payoff in the class of all direct admissible mechanisms or an indirect mechanism that implements the outcome of a front-runner pessimal direct mechanism. The following theorem describes a front-runner pessimal mechanism, which holds **F** down to his outside option of zero.

Theorem 5. *A front-runner pessimal mechanism asks **S** to report his block z, and asks **B** to report his type. Reporting **D** commits **B** to buying z at $p(0)$. Reporting **F** leads to no trade. In equilibrium, each trader reports truthfully.*

Proof. Reporting **D** commits **B** to trading amount z at $p(0)$, which is the best trade for **D**, but entails a negative payoff for **F**. Reporting **F** leads to no trade, which enables **F** to secure his outside

[21] **S** is not allowed to report a block exceeding his actual block. This restriction makes the set of **S**'s messages type-dependent. Nevertheless, the Revelation Principle applies because the Nested Range Condition of Green and Laffont (1986) holds.

option of zero, but is the worst trade for **D**. Therefore, each type of **B** prefers truthful reporting. Also **S** prefers truthful reporting because understating z would not reverse his no-trade with **F** and would reduce his trade with **D**. □

The outcome of the mechanism in Theorem 5 can be implemented by an indirect mechanism that uses iceberg orders and a restriction on minimal order sizes. In this mechanism, **S** may submit an iceberg sell order for any amount at price $p(0)$. **B** may submit an iceberg buy order only for an amount weakly exceeding $\bar{z}$ (the upper bound of the support of z). In equilibrium, **S** submits an iceberg order to sell his entire block z and specifies that only amount δ be visible. **D** responds with a limit order (i.e., an iceberg order with no hidden part) to buy amount $\bar{z}$. This order is executed immediately (and partially if $z < \bar{z}$). **F** does not respond to **S**'s order. The restriction on the minimal size of responding orders discourages **F** from probing **S**'s invisible quantity by submitting buy orders for smaller quantities in order to infer from their execution the likely magnitude of **S**'s order. **S**'s concealment of the exact size of his intended trade from **F** by using an iceberg order keeps **F**'s payoff low by keeping **F** uninformed.

4.2 A Surplus-Maximizing Mechanism

A **surplus-maximizing mechanism** is a direct mechanism that maximizes the surplus

$$\mathbb{E}_{|\mathbf{B}=\mathbf{F}}\left[-\mathbf{1}_{\{\mathbf{F}\text{ front-runs}\}}c - \mathbf{1}_{\{z>Q\}}r\right] \tag{9}$$

over the class of direct admissible mechanisms, or an indirect mechanism that implements the outcome of a surplus-maximizing direct mechanism. The expectation in (9) is with respect to **B**'s belief f, and Q is a random variable driven by z and denotes **S**'s trade with **B**.

Surplus-maximization may be sought by the exchange that, due to regulatory or competitive pressure, is concerned about the sum of traders' payoffs.[22] This sum is the surplus in (9) except the terms that are independent of the exchange's choice of a mechanism, provided the mechanism is such that **D** buys **S**'s entire block, which will be assumed and can be shown to entail no loss of

[22]The sum of traders' payoffs includes the payoffs of **S** and **B**, and the (hitherto suppressed) payoffs of the traders represented by the inverse demand function p, specified so that for **S**'s every trade, prices affect the distribution of payoffs, but not their sum.

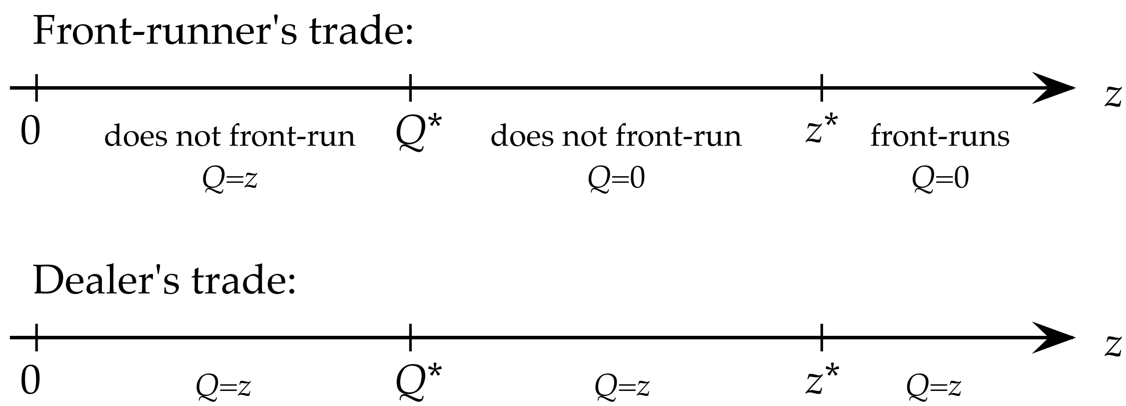

Figure 1: In the surplus-maximizing mechanism when $\delta \to 0$, **F** buys **S**'s entire block z provided it is sufficiently small, whereas **D** always buys **S**'s entire block. **B**'s trade is denoted by Q.

generality.[23]

The following theorem describes a surplus-maximizing direct mechanism, which Figure 1 illustrates. For tractability, the focus is on the case with $\delta \to 0$.

Theorem 6. *Let $\delta \to 0$, and let either π be sufficiently small or r be sufficiently large. A surplus-maximizing mechanism is characterized by thresholds Q^* and z^* satisfying $0 < Q^* \leq z^* < \bar{z}$. This mechanism asks* **S** *to report his block z, and asks* **B** *to report his type,* **D** *or* **F***. Reporting* **D** *commits* **B** *to buying z. Reporting* **F** *when $z \leq Q^*$ triggers a recommendation not to front-run and commits* **B** *to buying z at $p(0)$. Reporting* **F** *when $z \in (Q^*, z^*)$ triggers a recommendation not to front-run and leads to no trade. Reporting* **F** *when $z \geq z^*$ triggers a recommendation to front-run and leads to no trade.*

In equilibrium, each trader reports truthfully and **F** *abides by the exchange's recommendation regarding whether to front-run.* **F***'s equilibrium payoff equals his outside option, so the mechanism is also front-runner pessimal. Moreover, the exchange's policy of disclosing* **S***'s reported block to* **F** *before* **F** *makes his front-running decision is suboptimal if and only if $Q^* < z^*$.*

Proof. See Appendix. □

The intuition for the mechanism described in Theorem 6 is the following. The exchange wants **F** to buy **S**'s entire block z as often as possible to save on costs c and r. Because buying from **S** is costly for **F** (due to the price impact from the subsequent liquidation), **F** must be compensated

[23] The payoff of **D** is maximized if he buys **S**'s entire block, which can be ensured in any mechanism without tempting **F** to mimic **D**'s equilibrium strategy. Such a mechanism must merely promise that **B** who declares himself to be **D** must trade **S**'s entire block. The proof of Theorem 6 has the details.

for this purchase. This compensation comes as a profit from front-running. **F** benefits from front-running more when z is higher. The exchange's loss from front-running, c, does not depend on z. Therefore, when $z \geq z^*$, **F** is recommended to front-run, without having to buy anything from **S**. Moreover, the exchange's loss r from **S**'s unfulfilled trade does not depend on the size of that trade, and so the exchange prefers to sell smaller blocks to **F** first. Selling small blocks helps limit **F**'s losses from trading, thereby ensuring his participation. As a result, **F** trades when $z \leq Q^*$, but not when $Q^* \in (z, z^*)$.

The outcome of the described direct mechanism can be implemented by an indirect mechanism that uses iceberg orders, a restriction on minimal order sizes, and occasional flashing. The exchange **flashes** when it discloses an iceberg order's purportedly invisible part.[24] In the proposed indirect mechanism, **S** may submit an iceberg sell order for any amount. **B** has two options: (i) to submit an iceberg buy order for an amount weakly exceeding $\bar{z}$, or (ii) to agree to a transaction that flashes **S**'s iceberg order if and only if its size exceeds z^*, and makes **B** buy z at price $p(0)$ otherwise.

In equilibrium, **S** submits an iceberg order to trade his entire block z at price $p(0)$ and specifies that only amount δ be visible to others. (Because the assumption has been that $\delta \to 0$, the entire order is essentially invisible.) **D** chooses option (i), by submitting, for instance, a limit buy order for amount $\bar{z}$, which may be filled only partially (if $z < \bar{z}$). **F** chooses option (ii), and front-runs if and only if z is flashed.

The nature of the exchange's incentives to conceal **S**'s block from **F**—and hence the indispensability of iceberg orders in the proposed indirect mechanism—depends on whether $Q^* = z^*$ or $Q^* < z^*$. Either case is possible, depending on parameters. Suppose that $Q^* = z^*$. Then, **S**'s iceberg order can be replaced by a limit order. When $z \geq z^*$, **F** is recommended to front-run, and obeying by this recommendation remains optimal even after a limit order reveals z, as can be shown. The same logic applies when $z < Q^* = z^*$. By contrast, when $Q^* < z^*$ and z is close to, but less than, z^*, iceberg orders are indispensable. **F** is recommended not to front-run. If a limit order revealed z, however, **F** would disobey and front-run. **F** finds front-running unprofitable in

[24]In practice, exchanges sometimes flash submitted orders to selected traders. These flashes last for milliseconds and are believed to trigger front-running: "[...] the consumers of these short-duration, flashed quotes are, not surprisingly, often high-frequency shops, who can interpret them as imminent demand across the current spread and try to trade ahead," according to Michel Debiche, president and CEO of quantitative investment advisory firm Quantia Capital Management. (See http://www.thetradenews.com/asset-classes/equities/3353, accessed 27 January 2013.)

expectation only because an iceberg order conceals the exact value of z.

5 Conclusions

The paper's mechanism-design analysis supports the hypothesis that in markets in which workups are observed, the exchange is mistrusted, because iceberg orders can render the workup redundant when the exchange is trusted. Traditionally, workups were mediated by human brokers, who could trade both on behalf of their customers and for their own account, and thus could front-run easily. Today, some electronic exchanges may lack the reputation necessary to assure traders that these exchanges will not use the traders' information against them. If trust comes with reputation, one may further conjecture that expandable orders would be used in new exchanges, perhaps not yet trusted, whereas iceberg orders would be used in established exchanges, with a reputation for not abusing traders' private information.

The conclusion that the existence of the workup is indicative of the lack of trust in the exchange must be qualified. Iceberg orders used in practice feature the loss of time priority for the hidden part.[25] This feature may handicap iceberg orders relative to workups. An opposing practical consideration exists, however, that favors iceberg orders. In practice, the (unmodelled) delay caused by the workup may impose a negative externality on other traders by slowing down trade in the asset.[26]

The focus has been on the workup's equilibrium that minimizes the front-runner's expected payoff. It has been shown how the corresponding equilibrium outcome can be fully implemented in the button mechanism, a modification of the workup. Another workup's equilibrium that may appeal to the exchange is the one that maximizes the sum of the traders' expected payoffs.[27] Such an equilibrium—by contrast to the gradual-disclosure equilibrium—may have jumps (i.e., offers to increase the agreed-upon trade by an amount exceeding the smallest feasible increment). A jump convinces the front-runner that the seller's block is sufficiently large. The front-runner then may accept the jump in order to retain the possibility of receiving and making future offers, thereby

[25] The loss of time priority means that if a trader submits a visible order at the price of an existing invisible order, the visible order is executed first, even though it was submitted last.

[26] Garbade (1978) was the first to point out this externality.

[27] Theorem 6 describes a mechanism that maximizes the sum of the traders' expected payoffs. The theorem is mute about the workup's equilibrium that maximizes the sum of the traders' expected payoffs.

learning about the seller's block. Without the possibility of jumps, the front-runner may be unable to overcome his initial prior belief that the seller's block is small, and hence may not trade at all, to the detriment of the traders' surplus. Full exploration of such signalling equilibria is beyond this paper's scope and is related to the work of Hörner and Skrzypacz (2009).

References

Abreu, Dilip and Faruk Gul, "Bargaining and Reputation," *Econometrica*, 2000, *68* (1), 85–117.

Admati, Anat R. and Paul Pfleiderer, "Direct and Indirect Sale of Information," *Econometrica*, 1990, *4* (58), 901–928.

Admati, Anat Ruth and Paul Pfleiderer, "Sunshine Trading and Financial Market Equilibrium," *Review of Financial Studies*, 1991, *4* (3), 443–481.

Aitken, Michael J., Henk Berkman, and Derek Mak, "The use of Undisclosed Limit Orders on the Australian Stock Exchange," *Journal of Banking & Finance*, 2001, *25* (8), 1589–1603.

Anand, Amber and Daniel G. Weaver, "Can order exposure be mandated?," *Journal of Financial Markets*, 2004, *7* (4), 405–426.

Bernhardt, Dan and Bart Taub, "Front-running Dynamics," *Journal of Economic Theory*, 2008, *138* (1), 288–296.

Boni, Leslie and Chris Leach, "Expandable Limit Order Markets," *Journal of Financial Markets*, 2004, *7*, 145–185.

Brunnermeier, Markus K. and Lasse Heje Pedersen, "Predatory Trading," *The Journal of Finance*, 2005, *60* (4), 1825–1863.

Burdett, Kenneth and Maureen O'Hara, "Building Blocks: An Introduction to Block Trading," *Journal of Banking and Finance*, 1987, *11*, 193–212.

Buti, Sabrina and Barbara Rindi, "Hidden Orders and Optimal Submission Strategies in a Dynamic Limit Order Market," *Rotman School of Management Working Paper*, 2009.

Danthine, Jean-Pierre and Serge Moresi, "Front-Running by Mutual Fund Managers: A Mixed Bag," *European Finance Review*, 1998, *2* (1), 29–56.

D'Souza, Chris and Charles Gaa, "The Effect of Economic News on Bond Market Liquidity," *Bank of Canada Working Paper*, 2004, (16).

Duffie, Darrell, *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets (Princeton Lectures in Finance)*, Princeton University Press, 2011.

_ **and Bruno Strulovici**, "Capital Mobility and Asset Pricing," *Working Paper*, 2011.

Dungey, Mardi, Olan Henry, and Michael McKenzie, "From Trade-to-Trade in US Treasuries: Durations, Workups and News Effects," 2009.

Esser, Angelika and Burkart Mönch, "The navigation of an iceberg: The optimal use of hidden orders," *Finance Research Letters*, 2007, *4* (2), 68–81.

Evans, Robert, "Sequential Bargaining with Correlated Values," *Review of Economic Studies*, 1989, *56* (4), 499–510.

Fishman, Michael J and Francis A. Longstaff, "Dual Trading in Futures Markets," *Journal of Finance*, 1992, *47* (2), 643–671.

Fleming, Michael and Bruce Mizrach, "The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform," *Federal Reserve Bank of New York Staff Reports*, 2009, (381).

Fudenberg, Drew and Jean Tirole, *Game Theory*, The MIT Press, 1991.

Garbade, Kenneth D., "The Effect of Interdealer Brokerage on the Transactional Characteristics of Dealer Markets," *Journal of Business*, 1978, *51* (3), 477–498.

Green, Jerry and Jean-Jacques Laffont, "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 1986, *53* (3), 447–456.

Grossman, Sanford J., "An Economic Analysis of Dual Trading," *Rodney L. White Center for Financial Research Working Paper*, 1989.

Harris, Christopher and John Vickers, "Perfect Equilibrium in a Model of a Race," *Review of Economic Studies*, 1985, *52* (2), 193–209.

Harris, Larry, *Trading and Exchanges: Market Structure for Practitioners*, Oxford University Press, 2003.

Harris, Lawrence E., "Order Exposure and Parasitic Traders," *Marshall School of Business Working Paper*, 1997.

Hörner, Johannes and Andrzej Skrzypacz, "Selling Information," 2009.

Huang, Roger D., Jun Cai, and Xiaozu Wang, "Information-Based Trading in the Treasury Note Interdealer Broker Market," *Journal of Financial Intermediation*, 2002, *11*, 269–296.

Kastl, Jakub, "On the Properties of Equilibria in Private Value Divisible Good Auctions with Constrained Bidding," *Journal of Mathematical Economics*, 2012, *48* (6), 339–352.

Kremer, Ilan and Kjell G. Nyborg, "Underpricing and Market Power in Uniform Price Auctions," *The Review of Financial Studies*, 2004, *17* (3), 849–877.

Kuznetsov, Alex, *The Complete Guide to Capital Markets for Quantitative Professionals*, McGraw-Hill, 2006.

Kyle, Albert S, "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

Luenberger, David G., *Optimization by Vector Space Methods*, New York: John Wiley & Sons, 1969.

Madigan, B. and J. Stehm, "An Overview of the Secondary Market for U.S. Treasury Securities in London and Tokyo," *Federal Reserve Board Finance and Economics Discussion Paper*, 1994, pp. 94–117.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

Mizrach, Bruce and Christopher J. Neely, "The Transition to the Electronic Communications Networks in the Secondary Treasury Market," *Federal Reserve Bank of St. Louis Review*, November/December 2006, *88* (6), 527–541.

Moinas, Sophie, "Hidden limit orders and liquidity in order driven markets," *Toulouse School of Economics Working Paper*, 2010.

Myerson, Roger B., *Game Theory: Analysis of Conflict*, Harvard University Press, 1997.

Pagano, M. and A. Röell, *Financial Markets' Liberalization and the Role of Banks*, Cambridge: Cambridge University Press, 1993.

Rochet, Jean-Charles and Jean-Luc Vila, "Insider Trading without Normality," *The Review of Economic Studies*, 1994, *61* (1), 131–152.

Röell, Ailsa, "Dual-Capacity Trading and the Quality of the Market," *Journal of Financial Intermediation*, 1990, *1*, 105–124.

Schwartz, Robert A., Reto Francioni, and Bruce W. Weber, eds, *The Equity Trader Course*, John Wiley & Sons, Inc., 2006.

Vayanos, Dimitri, "Strategic Trading in a Dynamic Noisy Market," *The Journal of Finance*, 2001, *56* (1), 131–171.

Vincent, Daniel R., "Bargaining with Common Values," *Journal of Economic Theory*, 1989, *48* (1), 47–62.

Wilson, Robert, "Auctions of Shares," *The Quarterly Journal of Economics*, 1979, *93* (4), 675–689.

A Appendix: Omitted Proofs

Proof of Theorem 1: An **on-equilibrium-path history** is a history consistent with equilibrium play for some type profile. An **off-equilibrium-path history** is a history that is not on-equilibrium-path. A trader's **observable deviation** is an action that, given a history, is inconsistent with equilibrium strategy for the trader's every type that is in the support of the other trader's belief.

Traders' equilibrium strategies after on-equilibrium-path histories are described in the theorem's statement. After any off-equilibrium-path history,

- **S** rejects if the history has **B**'s observable deviation, and reverts to his on-equilibrium-path strategy otherwise.

- **F** rejects, liquidates, and front-runs.
- **D** offers δ and accepts any offer.

These equilibrium strategies are supported by traders' beliefs whose on-equilibrium-path components are derived from equilibrium strategies using Bayes's rule. These beliefs' off-equilibrium-path components are revised using Bayes's rule whenever possible, and when impossible, are assumed to be as follows:

- After **B**'s every observable deviation, **S** assigns probability one to **B** being **F**. Whenever **S**'s observable deviation is followed by **B**'s rejection, **S** assigns probability one to **B** being **F**; otherwise (i.e., if **B** fails to reject), **S** assigns probability one to **B** being **D**.
- Respecting condition B(i) of Fudenberg and Tirole (1991, Section 8.2.3), both types of **B** form beliefs according to the same rule. In particular, after **S**'s every observable deviation, **B** assigns probability one to $z = \bar{z}$.
- Respecting condition B(iii) of Fudenberg and Tirole (1991, Section 8.2.3), neither trader revises his belief about the other trader in response to his own deviations (observable or not).

The appeal to Bayes's rule whenever possible respects condition B(ii) of Fudenberg and Tirole (1991, Section 8.2.3).

The remainder of the proof verifies that the described strategy profile and the system of beliefs constitute the workup game's perfect Bayesian equilibrium defined to satisfy Definition 8.2 of Fudenberg and Tirole (1991, p. 333).[28]

1. ***S**'s strategy*:

   (a) *On equilibrium path*:

      i. Suppose $Q_{n-1} = z$. Then, feasibility requires that **S** reject.

      ii. Suppose $Q_{n-1} < z$. It will be shown that for a sufficiently small π or a sufficiently large r, at any round n after an on-equilibrium-path history, **S** (strictly) prefers offering δ or accepting $q_n = \delta$ (whichever is appropriate) to rejecting (an increase

[28]Here, condition B(iv) of Fudenberg and Tirole (1991, Section 8.2.3) is vacuous because there are fewer than three players.

proposed by **B** or by **S** himself, or the opportunity to suggest a further increase). This and subsequent claims in this proof use the one-stage-deviation principle of optimality, which consists in comparing each trader's payoff from his prescribed equilibrium strategy to his payoff from a round-n deviation followed by reverting to his equilibrium strategy.

If $Q_{n-1} \geq \hat{Q}$, **S** infers that **B** is **D**. In this case, **S**'s payoff from his equilibrium strategy exceeds his payoff from rejection: $p(0)\,z > p(0)\,Q_{n-1} + \int_0^{z-Q_{n-1}} p(s)\,\mathrm{d}s - r$.

Suppose $Q_{n-1} < \hat{Q}$. If **S** deviates by rejecting, his payoff is at most[29]

$$p(0)\,z - \left(r + \int_0^{z-Q_{n-1}} \left(p(0) - (1-\pi)\,p(s) - \pi p\left(Q_{n-1}+s\right)\right)\mathrm{d}s\right), \tag{10}$$

which assumes that **F** does not front-run.[30] If, by contrast, **S** adheres to his prescribed equilibrium strategy, his expected payoff is

$$p(0)\,z - \mathbf{1}_{\left\{z \geq \hat{Q}\right\}}\pi\left(r + \int_0^{z-\left(\hat{Q}-\delta\right)} \left(p(0) - p\left(\hat{Q}-\delta+k+s\right)\right)\mathrm{d}s\right), \tag{11}$$

which uses the fact that, if $z \geq \hat{Q}$, **F** front-runs.

By inspection of (10) and (11), $z < \hat{Q}$ immediately implies that **S** strictly prefers not to reject. Instead, suppose $z \geq \hat{Q}$. Then, **S** prefers not to reject if (11) exceeds (10), which occurs whenever

$$\begin{aligned} r(1-\pi) > \pi \int_0^{z-\left(\hat{Q}-\delta\right)} &\left(p(0) - p\left(\hat{Q}-\delta+k+s\right)\right)\mathrm{d}s \\ &- \int_0^{z-Q_{n-1}} \left(p(0) - (1-\pi)\,p(s) - \pi p\left(Q_{n-1}+s\right)\right)\mathrm{d}s. \end{aligned} \tag{12}$$

By inspection, (12) holds if r is sufficiently large. Formally, for all π, p, z, k, Q_{n-1}, and $\hat{Q}$, there exists $\bar{r}$ such that for all δ such that Q_{n-1} and $\hat{Q}$ are multiples of δ and

[29] If **S** rejects his own offer, **F** front-runs in response to **S**'s observable deviation. In this case, **S**'s payoff is lower than indicated by (10). If **S** rejects **B**'s offer, **S**'s deviation is not observable, **F** (wrongly) infers that $z = Q_{n-1}$, and does not front-run.

[30] The integral in (10) is the cumulative discount on the $z - Q_{n-1}$ units that **S** chooses not to try to sell in the workup. This discount captures the price impact.

for all $r > \bar{r}$, (12) holds. Furthermore, (12) holds also if π is sufficiently small. Indeed, each integral in (12) is positive because $z > \hat{Q} - \delta \geq Q_{n-1}$ and $\hat{Q} \geq \delta$). Hence, as $\pi \to 0$, the product of π and the first integral vanishes, and the right-hand side of (12) converges to a negative number, whereas the left-hand side converges to a positive r.

iii. Continue to suppose $Q_{n-1} < z$. It will be shown that, at any round n after an on-equilibrium-path history, **S** weakly prefers offering $q_n = \delta$ to offering $q_n \neq \delta$, and that this preference is strict if either **F** follows type-(i) strategy and $n = 1$ or **F** follows type-(ii) strategy and $Q_{n-1} < \hat{Q} - \delta$.[31]

If **F** follows type-(i) strategy, **S**'s equilibrium payoff $(1-\pi)\, p\,(0)\, z + \pi \left(-r + \int_0^z p\,(s)\, \mathrm{d}s\right)$ exceeds his payoff from offering $q_n \neq \delta$ and triggering **F**'s front-running: $(1-\pi)\, p\,(0)\, z + \pi \left(-r + \int_0^z p\,(k+s)\, \mathrm{d}s\right)$. Henceforth suppose **F** follows type-(ii) strategy with a threshold $\hat{Q}$.

If the traders have committed to $Q_{n-1} \geq \hat{Q}$, **S** infers that **B** is **D**, who will accept any offer. In this case, **S** is indifferent between offering $q_n = \delta$ and offering $q_n \neq \delta$.

If the traders have committed to $Q_{n-1} = \hat{Q} - \delta$, **S** believes that **F** will front-run any offer q_n, and **D** will accept any offer q_n. Then, again, **S** is indifferent between offering $q_n = \delta$ and $q_n \neq \delta$.

If the traders have committed to $Q_{n-1} < \hat{Q} - \delta$, **S** believes that **F** will reject and front-run any offer $q_n \neq \delta$ and will accept $q_n = \delta$, and that **D** will accept any offer q_n. If **S** deviates by offering $q_n \neq \delta$, his payoff is

$$(1-\pi)\, p\,(0)\, z + \pi \left(p\,(0)\, Q_{n-1} + \int_0^{z-Q_{n-1}} p\,(Q_{n-1} + k + s)\, \mathrm{d}s - r \right). \qquad (13)$$

If, by contrast, **S** adheres to his prescribed equilibrium strategy, his expected payoff

[31] These two types of the strategies are described in the theorem's statement.

is

$$
\begin{aligned}
&(1-\pi)\,p(0)\,z \\
&+\pi\left(\mathbf{1}_{\{z<\hat{Q}\}}\,p(0)\,z+\mathbf{1}_{\{z\geq\hat{Q}\}}\left(p(0)\left(\hat{Q}-\delta\right)+\int_{0}^{z-\left(\hat{Q}-\delta\right)} p\left(\hat{Q}-\delta+k+s\right)\mathrm{d}s-r\right)\right).
\end{aligned} \tag{14}
$$

S prefers offering $q_n = \delta$ to offering $q_n \neq \delta$ whenever the difference between (14) and (13) is positive:

$$
\pi \int_{Q_{n-1}}^{\min\left\{z,\hat{Q}-\delta\right\}} \left(p(0)-p(k+s)\right)\mathrm{d}s+\mathbf{1}_{\{z<\hat{Q}\}}\,\pi r>0. \tag{15}
$$

Indeed, inequality (15) follows from $Q_{n-1} < \hat{Q}-\delta$ and $Q_{n-1} < z$ (as has been assumed).[32]

(b) *Off equilibrium path*:

i. Consider an off-equilibrium-path history terminating in **S**'s (deviant) rejection of his own offer. This rejection terminates the workup; **S** can entertain no further actions.

ii. Consider an off-equilibrium-path history terminating in **S**'s (deviant) offer of $q_n \neq \delta$. If **B** is **F**, then **B** rejects (and liquidates and front-runs); **S** can entertain no further actions. If **B** is **D**, then **B** accepts, in which case **S** assigns probability one to **B** being **D**, and optimally reverts to his on-equilibrium-path strategy.

iii. Consider an off-equilibrium-path history terminating in **B**'s (deviant) offer $q_n \neq \delta$, leading **S** to believe that **B** is **F**. If **S** accepts (deviating from his equilibrium strategy), **F** responds to this deviation by rejecting, liquidating, and front-running. Hence, **S** can do no better than to reject **B**'s offer $q_n \neq \delta$.

iv. Consider an off-equilibrium-path history terminating in **B**'s rejection before **S** has revealed amount $\hat{Q}$. This rejection terminates the workup; **S** can entertain no fur-

[32]When $z < \hat{Q}$, the left-hand side of (15) captures a deviating **S**'s expected loss due to front-running that could have been avoided. When $z \geq \hat{Q}$, the left-hand side of (15) captures a deviating **S**'s loss due to front-running that could not have been avoided, but could have been postponed.

ther actions.

2. ***D*'s strategy on and off the equilibrium path**:

   (a) Because **D**'s payoff is increasing in his trade with **S**, it is a weakly dominant strategy for **D** to never reject.

   (b) If **D** proposes $q_n \neq \delta$, **S** rejects. Because **D**'s payoff is increasing in his trade with **S**, it is a weakly dominant strategy for **D** to propose $q_n = \delta$.

3. ***F*'s strategy**:

   (a) *On equilibrium path:*

      i. **F** never liquidates (with or without front-running) immediately after having accepted an offer, because a higher payoff can be attained by rejecting the offer and then liquidating (with or without front-running). Thus, **F** liquidates only after he or **S** rejects, or after **S** makes or accepts an offer.

      ii. **F** weakly prefers rejecting to offering $q_n \neq \delta$. Indeed, if **F** offers $q_n \neq \delta$, **S** rejects, leading to the same payoff that **F** could have obtained by rejecting himself.

      iii. It will be shown that if **F**'s type-(i) strategy is suboptimal, his type-(ii) strategy is optimal. Suppose that type-(i) strategy is suboptimal. That is, it is suboptimal for **F** to reject **S**'s round-1 offer and to liquidate without front-running. For any on-equilibrium-path history, **F**'s strategy can be represented as a mapping from the amount of the asset just revealed by **S** into the set of three elements: (a) reject, liquidate, and not front-run, (b) reject, liquidate, and front-run, and (c) accept and either offer δ (if it is **F**'s turn to propose) or accept (if it is **F**'s turn to propose). (Recall that it is suboptimal for **F** to liquidate immediately after having accepted, and that offering anything other than δ is equivalent to either (a) or (b).) Note that taking action (a) at round 1 constitutes type (i) strategy.

      If **S**'s latest action has been rejection, **F** infers that $Q_{n-1} = z$. Hence, **F** cannot profit by front-running; action (b) is suboptimal. Action (c) is infeasible. Hence, action (a) is optimal.

If **S**'s latest action has been to offer δ or to accept **F**'s offer of δ, it is suboptimal for **F** to take action (a). Indeed, taking action (a) at round $n \geq 2$ is dominated by taking action (a) at round 1, which corresponds to adopting type-(i) strategy, which is suboptimal by hypothesis. (The early exercise of action (a) enables **F** to avoid buying anything from **S** and selling with the price impact.)

Therefore, if **S**'s latest action has been to offer δ, it is optimal for **F** to take either action (b) or (c). This strategy can be captured by a real number, denoted by $\hat{Q}$. Once the amount revealed by **S** reaches $\hat{Q}$, **F** takes action (b). Until then, **F** takes action (c). Thus emerges type-(ii) strategy.

(b) *Off equilibrium path:*

i. Consider an off-equilibrium-path history terminating in **S** offering $q_n \neq \delta$, or rejecting his own offer, or rejecting at round 1. Then **F** believes $z = \bar{z}$. Given this belief, **F** rejects, liquidates, and front-runs.

ii. Consider an off-equilibrium-path history terminating in **F**'s (deviant) offer $q_n \neq \delta$. **S** rejects; F has no further actions to take.

Proof of Theorem 2: Let $q \equiv [q_n(z)]_{z,n}$ denote an on-equilibrium-path offer profile at some equilibrium of the workup, where $q_n(z)$ is the additional amount of **S**'s block revealed to **B** at round n when **S**'s block (invisible to **B**) is z. Let $m \equiv [m_n(z)]_{z,n}$ denote a sequence of *ad-hoc* announcements, where a multiset $m_n(z)$ is an announcement that **B** receives in the beginning of round n when **S**'s block is z.[33]

The proof proceeds in three steps, in the course of which the offer profile q is sequentially modified without requiring that any but the final modified profile be an on-path offer profile for some equilibrium. *Ad-hoc* announcements will be added, then removed. The focus is on changes in **F**'s payoffs along a sequence of the workup-induced fictitious decision problems in which **F** takes the revealed offer profile and the announcement sequence as given and, in every round, decides whether to simply liquidate or to liquidate and front-run.

Step 1 Take an offer profile q and an announcement sequence m. Suppose that q has a jump at

[33]For example, $m_n(z) = \{\delta, 2\delta, 0, 2\delta, \delta\}$ is a possible announcement; that is, an announcement can contain elements that appear more than once.

round n^*; that is, $q_{n^*}(z) > \delta$ for some z. At the beginning of round n^*, announce $q_{n^*}(z)$ to **F**, thus creating a new announcement sequence m', and spread the round-n^* jump over two rounds, thus creating a new offer profile q'—for every $z \in Z$:[34]

$$\begin{aligned}
q'_n(z) &= q_n(z) \quad \text{and} \quad m'_n(z) = m_n(z) \quad \text{if } n < n^* \qquad (16)\\
q'_{n^*}(z) &= \min\{\delta, q_{n^*}(z)\} \quad \text{and} \quad m'_{n^*}(z) = q_{n^*}(z)\\
q'_{n^*+1}(z) &= \max\{0, q_{n^*}(z) - \delta\} \quad \text{and} \quad m'_{n^*+1}(z) = \varnothing\\
q'_n(z) &= q_{n-1}(z) \quad \text{and} \quad m'_n(z) = m_{n-1}(z) \quad \text{if } n > n^* + 1.
\end{aligned}$$

Note the following **belief equivalence 1**:

- at any round $n \leq n^*$, (q, m) and (q', m') both lead **F** to hold the same posterior belief about z;
- at round $n^* + 1$, given (q', m'), **F** holds the same posterior belief about z as he does at n^*;
- at rounds $n^* + 2$, $n^* + 3$, etc., given (q', m'), **F** holds the same posterior belief about z as he does at rounds, respectively, $n^* + 1$, $n^* + 2$, etc., given (q, m).

It will be shown that **F**'s payoff with (q', m') is the same as with (q, m). Fix z. Given (q', m'), any strategy that involves accepting at n^* and liquidating (with or without front-running) at $n^* + 1$ is dominated by advancing the round-$(n^* + 1)$ action to round n^*, because doing so avoids commitment to a costly purchase of an extra δ. Then, display (16) and the belief equivalence 1 just described imply a one-to-one payoff correspondence between **F**'s actions with (q, m) and with (q', m') once the dominated strategy has been disregarded. In particular, the payoff from liquidating with or without front-running at round $n \in \{1, 2, .., n^*, n^* + 2, n^* + 3, ...\}$ with (q', m') is the same as the payoff from taking the same action at round $n - \mathbf{1}_{\{n > n^*\}}$ with (q, m). Hence, **F**'s payoffs from optimal strategies with (q, m) and (q', m') are the same.

The iteration of the procedure described in this step produces a new profile that has no jumps and, accompanied by additional announcements, gives the same expected payoff to **F** as the original profile, which has no announcements.

[34]Operators min and max in display (16) are necessary to accommodate z for which $q_{n^*}(z) = 0$. Recall that the assumption is that $q_{n^*}(z) > \delta$ for some z, not all.

Step 2 Take an offer profile q and an announcement sequence m. Suppose that q has a gap at round n^*; that is, $q_{n^*}(z) = 0$ for some z. For this z, at the beginning of round n^*, announce $q_{n^*}(z)$ to **F**, thus creating a new announcement sequence m', and close the round-n^* gap at n^*, thus creating a new offer profile q'. For z with no gap at n^*, leave the offer profile and the announcement sequence unchanged. Thus, given n^*, for any $z \in Z$:

$$q_{n^*}(z) \neq 0 \implies q'_n(z) = q_n(z) \quad \text{and} \quad m'_n(z) = m_n(z) \quad \text{for all } n \tag{17}$$
$$q_{n^*}(z) = 0 \implies \begin{cases} q'_n(z) = q_n(z) \quad \text{and} \quad m'_n(z) = m_n(z) \quad \text{if } n < n^* \\ q'_{n^*}(z) = q_{n+1}(z) \quad \text{and} \quad m'_{n^*}(z) = m_{n^*} \uplus m_{n^*+1}(z) \uplus q_{n^*}(z) \\ q'_n(z) = q_{n+1}(z) \quad \text{and} \quad m'_n(z) = m_{n+1}(z) \text{ if } n > n^*, \end{cases}$$

where $\uplus$ is the multiset sum operator, which ensures that an announcement can accommodate multiple identical messages (e.g., when multiple consecutive gaps are filled as this step is applied iteratively).

Note the following **belief-equivalence 2**:

- at any round $n < n^*$, (q, m) and (q', m') both lead **F** to hold the same posterior belief about z;
- $q_{n^*}(z) \neq 0$ implies that (q, m) and (q', m') both lead **F** to hold the same posterior belief about z at any round;
- $q_{n^*}(z) = 0$ implies that, at rounds n^*, $n^* + 1$, etc., given (q', m'), **F** holds the same posterior belief about z as he does at rounds, respectively, $n^* + 1$, $n^* + 2$, etc., given (q, m).

It will be shown that **F**'s payoff with (q', m') is the same as with (q, m), even though **F**'s information set corresponding to round n^* with $q_{n^*}(z) = 0$ and given (q, m) has no counterpart with (q', m'). Given (q, m), any strategy that involves liquidating (with or without front-running) at n^* is weakly dominated by postponing this action until $n^* + 1$ by accepting at n^*. Then, display (17) and the belief equivalence 2 just described imply a one-to-one payoff correspondence between **F**'s actions with (q, m) and with (q', m'), once the weakly dominated strategy has been disregarded. Hence, **F**'s payoffs from optimal strategies with (q, m) and (q', m') are the same.

The iteration of the procedure described in this step produces a new profile that has no jumps and, accompanied by additional announcements, gives the same expected payoff to **F** as the profile inherited from Step 1.

Step 3 By Steps 1 and 2, from any offer profile, one can construct a gradual offer profile (i.e., a profile with no gaps and no jumps) that has announcements and that gives **F** the same expected payoff as the original profile. Now, remove the sequence of announcements, denoted by m in Steps 1 and 2. As a result, **F** is weakly worse off because, in his decision problem of whether and when to front-run, he now has less information. The gradual offer profile can be sustained in an equilibrium of the workup game by Theorem 1. Hence, the gradual-disclosure mechanism is front-runner pessimal.

Proof of Theorem 6: Let the mechanism force **B** who reports **D** buy **S**'s entire block. Then, **D** will optimally report **D**, as doing so gives him the highest payoff he can hope for.

The mere requirement that the mechanism be voluntary ensures that **F** prefers to report his type truthfully. Whatever truthful reporting entails, it cannot give a negative payoff, because the mechanism is voluntary. By contrast, reporting **D** would lead to a negative payoff $-\int_0^z [p(0) - p(s)]\, \mathrm{d}s$ for **F**.

The surplus-maximizing mechanism's remaining features will be found by solving a certain relaxed problem. It will be shown that when r is large or π is small, a surplus-maximizing mechanism solves the following **relaxed problem**:

$$\max_{\phi,Q} \int_0^{\bar{z}} \left(-c\phi(z) - r\mathbf{1}_{\{Q(z)<z\}}\right) f(z)\, \mathrm{d}z \qquad \text{s.t.} \tag{18}$$

$$\int_0^{\bar{z}} \left(\phi(z)\left(\int_0^k [p(Q(z)+s) - p(z+s)]\, \mathrm{d}s - c\right) - \int_0^{Q(z)} [p(0) - p(s)]\, \mathrm{d}s\right) f(z)\, \mathrm{d}z \geq \bar{U}, \tag{19}$$

where, for each z, $\phi(z) = 1$ if **F** front-runs and $\phi(z) = 0$ if **F** liquidates, where $Q(z) \in [0, z]$ is **F**'s trade with **S**, and where

$$\bar{U} \equiv \max\left\{0, \int_0^{\bar{z}} \left(\int_0^k [p(s) - p(z+s)]\, \mathrm{d}s - c\right) f(z)\, \mathrm{d}z\right\} \tag{20}$$

is **F**'s outside option from not participating and taking an optimal front-running decision. The objective function in (18) is the surplus when **B** is **F**; when **B** is **D**, the surplus is independent of the choice of the mechanism because **S**'s entire block is always traded. Constraint (19) requires that the mechanism be voluntary; this constraint ensures that **F**'s expected payoff from front-running less his expected cost of trading with **S** exceeds his outside option. What makes the problem relaxed is the omission of the constraints ensuring that **F** front-runs if and only if the exchange recommends doing so, and the omission of the constraint that **S** finds it optimal to truthfully report his block.

At the solution to the relaxed problem:

- $Q(z) \in \{0, z\}$. Indeed, if $Q(z) \in (0, z)$, then setting $Q(z) = 0$ does not change the value of the objective function, but relaxes (19).
- $Q(z) = z \implies \phi(z) = 0$. Indeed, if $Q(z) = z$, then setting $\phi(z) = 0$ increases the value of the objective function but relaxes (19).
- Constraint (19) binds. Indeed, if the inequality in (19) were strict, the value of the objective function would be increased by enlarging the set of z for which $Q(z) = z$ and $\phi(z) = 0$ without affecting the strictness of the inequality.

Let λ be the Lagrange multiplier on (19). Then, the Lagrangian for the relaxed problem is

$$\mathcal{L} = \int_0^{\bar{z}} \left[\begin{array}{c} \phi(z)\left(\lambda\left(\int_0^k [p(Q(z)+s) - p(z+s)]\,\mathrm{d}s - c\right) - c\right) \\ -\lambda \int_0^{Q(z)} [p(0) - p(s)]\,\mathrm{d}s - r\mathbf{1}_{\{Q(z)<z\}} \end{array} \right] f(z)\,\mathrm{d}z - \lambda \bar{U}, \tag{21}$$

where $\lambda > 0$ because (19) binds. To solve the relaxed problem, a saddle point of (21) will be identified.

Performing pointwise maximization of (21) using $Q(z) = z \implies \phi(z) = 0$ implies

$$Q(z) = z \iff r \geq \max\left\{0, \lambda\left(\int_0^k [p(s) - p(z+s)]\,\mathrm{d}s - c\right) - c\right\} + \lambda \int_0^z [p(0) - p(s)]\,\mathrm{d}s. \tag{22}$$

Because the right-hand side of the inequality in (22) is continuous and increasing in z,

$$Q(z) = z \iff z \leq Q^*$$

for some $Q^* \in (0, \bar{z})$, where $Q^* > 0$ follows by inspection of (22) and $Q^* < \bar{z}$ is established by contradiction. If $Q^* = \bar{z}$, the Lagrangian $\mathcal{L} = -\lambda \int_0^{\bar{z}} \int_0^z (p(0) - p(s))\, \mathrm{d}s f(z)\, \mathrm{d}z - \lambda \bar{U}$ can be made arbitrarily small by taking λ to be arbitrarily large; no saddle point with $Q^* = \bar{z}$ exists.

Suppose that Q^* is such that

$$\lambda \left(\int_0^k [p(s) - p(Q^* + s)]\, \mathrm{d}s - c \right) - c \geq 0. \tag{23}$$

Then, Q^* solves

$$r = \lambda \left(\int_0^k [p(s) - p(Q^* + s)]\, \mathrm{d}s - c \right) - c + \lambda \int_0^{Q^*} (p(0) - p(s))\, \mathrm{d}s, \tag{24}$$

and $\phi(z) = \mathbf{1}_{\{z \geq z^*\}}$, where $z^* = Q^*$.

Suppose that Q^* is such that (23) is violated. Then, Q^* solves

$$r = \lambda \int_0^{Q^*} [p(0) - p(s)]\, \mathrm{d}s, \tag{25}$$

and $\phi(z) = \mathbf{1}_{\{z \geq z^*\}}$, where z^* satisfies $z^* > Q^*$ and solves

$$\int_0^k [p(s) - p(z^* + s)]\, \mathrm{d}s - \left(1 + \frac{1}{\lambda}\right) c = 0. \tag{26}$$

Furthermore, $Q^* > 0$ and (19) imply $z^* < \bar{z}$. Moreover, $z^* > \hat{z}$, where $\hat{z}$ is the block size at which **F** is indifferent between front-running and not front-running:

$$\int_0^k [p(s) - p(\hat{z} + s)]\, \mathrm{d}s - c = 0.$$

Thus, if $z \in (Q^*, z^*) \cap (\hat{z}, z^*)$, the exchange asks **F** not to front-run (provided the saddle point of the relaxed problem contains a solution of the full problem, as will be shown), whereas **F** would front-run if he knew z. Thus, $Q^* < z^*$ implies that the exchange's policy of disclosing **S**'s block to **F** before **F** makes his front-running decision is suboptimal.

To summarize, the saddle point of the Lagrangian (21) is

$$Q(z) = \mathbf{1}_{\{z \le Q^*\}} z \text{ and } \phi(z) = \mathbf{1}_{\{z \ge z^*\}}, \text{ where} \tag{27}$$

$$\begin{cases} z^* = Q^* & \text{and} \quad Q^* \text{ solves (24) if (23) holds} \\ z^* > Q^* & \text{and} \quad Q^* \text{ solves (25) and } z^* \text{ solves (26) if (23) fails} \end{cases}$$

$$\lambda \text{ solves (19) that holds as equality.}$$

The sufficiency Theorem 2 (p. 221) of Luenberger (1969) implies that the saddle point's allocation component (Q, ϕ) solves the designer's relaxed problem.

It remains to perform verification, ascertaining that the solution to the relaxed problem is the solution to the exchange's problem.

To verify **F**'s incentives to abide by the exchange's recommendations, suppose first that the exchange recommends front-running. **F** infers that $z \ge z^*$ and $Q(z) = 0$, and obeys if

$$\int_{z^*}^{\bar{z}} \left(\int_0^k [p(s) - p(z+s)] \, ds - c \right) f(z) \, dz \ge 0,$$

which holds by (19) with the allocation component of the saddle point (27) substituted:

$$-\int_0^{Q^*} \int_0^z [p(0) - p(s)] \, ds f(z) \, dz + \int_{z^*}^{\bar{z}} \left(\int_0^k [p(s) - p(z+s)] \, ds - c \right) f(z) \, dz \ge \bar{U}. \tag{28}$$

Suppose that the exchange does not recommend front-running. If $z^* = Q^*$, **F** infers that $Q(z) = z$ and obeys trivially. If $z^* > Q^*$, **F** infers $z < z^*$ and obeys if (after some rearrangements)

$$-cF(Q^*) \, dz + \int_{Q^*}^{z^*} \left(\int_0^k [p(s) - p(z+s)] \, ds - c \right) f(z) \, dz \le 0. \tag{29}$$

Substituting the definition of $\bar{U}$ from (20) into the simplified voluntary participation constraint

(28) and rearranging gives

$$\int_{Q^*}^{z^*} \left(\int_0^k \left[p\left(s\right) - p\left(z+s\right) \right] \mathrm{d}s - c \right) f\left(z\right) \mathrm{d}z$$
$$\leq - \int_0^{Q^*} \int_0^z \left[p\left(0\right) - p\left(s\right) \right] \mathrm{d}s f\left(z\right) \mathrm{d}z - \int_0^{Q^*} \left(\int_0^k \left[p\left(s\right) - p\left(z+s\right) \right] \mathrm{d}s - c \right) f\left(z\right) \mathrm{d}z. \quad (30)$$

Then,

$$-cF\left(Q^*\right) + \int_{Q^*}^{z^*} \left(\int_0^k \left[p\left(s\right) - p\left(z+s\right) \right] \mathrm{d}s - c \right) f\left(z\right) \mathrm{d}z$$
$$\leq - \int_0^{Q^*} \int_0^z \left[p\left(0\right) - p\left(s\right) \right] \mathrm{d}s f\left(z\right) \mathrm{d}z - \int_0^{Q^*} \int_0^k \left[p\left(s\right) - p\left(z+s\right) \right] \mathrm{d}s f\left(z\right) \mathrm{d}z < 0,$$

where the first inequality is by (30), and the second inequality is immediate because each double integral is positive. The two inequalities combined imply the obedience constraint (29).

It remains to verify that **S** finds it optimal to truthfully report his block z:

1. Consider **S** of type $z \leq Q^*$. His payoff from reporting truthfully is maximal possible, and hence he would not profit from lying.

2. Consider **S** of type $z \in (Q^*, z^*)$. His payoff from reporting truthfully is

$$(1-\pi)\, z p\left(0\right) + \pi \left(\int_0^z p\left(s\right) \mathrm{d}s - r \right).$$

   (a) No lie $z' \geq z^*$ is feasible.

   (b) Every lie $z' \in (Q^*, z)$ is unprofitable because it does not affect **S**'s payoff.

   (c) The payoff from a lie $z' \leq Q^*$ is

$$z' p\left(0\right) - r + (1-\pi) \int_0^{z-z'} p\left(s\right) \mathrm{d}s + \pi \int_0^{z-z'} p\left(z'+s\right) \mathrm{d}s.$$

   This lie is unprofitable if

$$(1-\pi) \left(\int_0^{z-z'} \left(p\left(0\right) - p\left(s\right)\right) \mathrm{d}s + r \right) \geq \pi \int_0^{z'} \left(p\left(0\right) - p\left(s\right)\right) \mathrm{d}s,$$

which holds for a sufficiently small π or a sufficiently large r.

3. Consider **S** of type $z \geq z^*$. His payoff from reporting truthfully is

$$(1-\pi)\, zp\,(0) + \pi \left(\int_0^z p\,(k+s)\, \mathrm{d}s - r \right).$$

(a) The payoff from a lie $z' \in (Q^*, z^*)$ is

$$(1-\pi) \left(z'p\,(0) + \int_0^{z-z'} p\,(s)\, \mathrm{d}s \right) + \pi \int_0^z p\,(s)\, \mathrm{d}s - r.$$

This lie is unprofitable if

$$(1-\pi) \left(\int_0^{z-z'} [p\,(0) - p\,(s)]\, \mathrm{d}s + r \right) \geq \pi \int_0^z [p\,(s) - p\,(k+s)]\, \mathrm{d}s,$$

which holds for a sufficiently small π or a sufficiently large r.

(b) The payoff from a lie $z' \leq Q^*$ is

$$p\,(0)\, z' - r + \pi \int_0^{z-z'} p\left(z'+s\right) \mathrm{d}s + (1-\pi) \int_0^{z-z'} p\,(s)\, \mathrm{d}s.$$

This lie is unprofitable if

$$(1-\pi) \left(\int_0^{z-z'} (p\,(0) - p\,(s))\, \mathrm{d}s + r \right) \geq \pi \left(\int_0^{z'} (p\,(0) - p\,(k+s))\, \mathrm{d}s + \int_{z'}^{z} (p\,(s) - p\,(k+s))\, \mathrm{d}s \right),$$

which holds for a sufficiently small π or a sufficiently large r.