Quick or cheap? Breaking points in dynamic markets

Mertikopoulos, Panayotis ; Nax, Heinrich H ; Pradelski, Bary S R

DOI: https://doi.org/10.1016/j.jmateco.2024.102987

Journal of Mathematical Economics 112 (2024) 102987

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Quick or cheap? Breaking points in dynamic markets

Panayotis Mertikopoulos [a], Heinrich H. Nax [b], Bary S.R. Pradelski [c,*]

[a] Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France
[b] Univ. Zurich, 8050 Zurich, Switzerland
[c] CNRS, Maison Française d'Oxford, Oxford, OX2 6SE, United Kingdom

ARTICLE INFO

Manuscript handled by Editor S Takahashi

Keywords:
Dynamic matching
Online markets
Market design

ABSTRACT

We examine two-sided markets where players arrive stochastically over time. The cost of matching a client and provider is heterogeneous, and the distribution of costs – but not their realization – is known. In this way, a social planner is faced with two contending objectives:(*a*) to reduce the players' *waiting time* before getting matched; and (*b*) to reduce *matching costs*. In this paper, we aim to understand when and how these objectives are incompatible. We identify two regimes dependent on the 'speed of improvement' of the cost of matching with respect to market size. One regime results in a *quick or cheap* dilemma without 'free lunch': there exists no clearing schedule that is simultaneously optimal along both objectives. In that regime, we identify a unique breaking point signifying a stark reduction in matching cost contrasted by an increase in waiting time. The other regime features a window of opportunity in which free lunch can be achieved. Which scheduling policy is optimal depends on the heterogeneity of match costs. Under limited heterogeneity, e.g., when there is a finite number of possible match costs, greedy scheduling is approximately optimal, in line with the related literature. However, with more heterogeneity greedy scheduling is never optimal. Finally, we analyze a particular model where match costs are exponentially distributed and show that it is at the boundary of the no-free-lunch regime We then characterize the optimal clearing schedule for varying social planner desiderata.

1. Introduction

Many market interactions require the dynamic matching of heterogeneous agents that arrive stochastically to a two-sided market. Examples include the dynamic matching of clients and providers in markets for jobs and services, of buyers and sellers in financial markets, of taxis and passengers on road networks, of donors and recipients in organ exchanges, etc.

It is known that many of these markets vary substantially in terms of achieving desired outcomes (Roth and Xing, 1994, 1997). The focus of our investigation is on a crucial aspect of market design in this context, namely the *scheduling of clearing events*. The goal is to find the best schedule of market clearing so that sufficient clients and providers are in the market to allow for desirable matches over time while not waiting excessively. Designing an optimal clearing policy thus requires balancing the following two objectives:

1. To reduce the coexistence of agents on the two sides of the market.
2. To allow parties to match in such a way so as to minimize cost (or maximize productivity).

In pursuit of these two goals, clearing schedules need to be formulated to address the following key question: *How long should the social planner wait between two clearing events?*

As many market places commit to a clearing schedule and do not interfere in who trades with whom, we shall focus on the latter question and assume that who trades with whom is governed by a pre-defined mechanism. Thus, the social planner is left to only decide when to open the market place for matching events to take place, dependent on the number of players on each side of the market.

To illustrate the above, consider the example of a governmental employment bureau faced with a dynamically evolving job market where job offers and job seekers arrive to the system stochastically over time. Suppose that the bureau has to decide how often to send the list of candidates to firms and vice versa. The bureau thus has a single variable (when to send the lists) that it employs to balance between the coexistence of vacancies and job seekers, and to enable matches between vacancies with the skills of individual job seekers so as to maximize productivity. Waiting times incur costs via unemployment benefits, as well as costs due to productivity losses incurred by badly staffed vacancies.

* Corresponding author.
E-mail addresses: panayotis.mertikopoulos@imag.fr (P. Mertikopoulos), heinrich.nax@uzh.ch (H.H. Nax), bary.pradelski@cnrs.fr (B.S.R. Pradelski).

https://doi.org/10.1016/j.jmateco.2024.102987
Received 29 August 2023; Received in revised form 7 March 2024; Accepted 10 April 2024
Available online 17 April 2024
0304-4068/

The goal of our paper is to characterize the optimal scheduling of clearing events as a function of the distribution of random arrivals and their associated costs. To do this in as general a setting as possible, we take an application-agnostic approach and abstract away any application-specific details (such as the particular structure of the application and recruitment processes). Further, we do not consider strategic incentives for market participants to misreport their type (in the non-transferable utility context) or to bid strategically (in the transferable utility context). This allows us to focus on the trade-offs that arise between two different and concurrent objectives, waiting time versus matching cost. Perhaps surprisingly, this *'quick or cheap'* dilemma is often impossible to resolve. Even if there is no dilemma, greedy scheduling policies are generally not optimal in this context when match costs are fully heterogeneous. However, for limited heterogeneity we find that greedy scheduling is approximately optimal, thus providing a solid footing for several prior results in the literature.

Related work

Dating back to the 1950s, the first related strand of work focuses on behavioral aspects underlying the dynamics of unemployment and job vacancies in labor markets (Dow and Dicks-Mireaux, 1958). These analyses identify avenues to reduce *waiting* – i.e., the coexistence of unemployment and vacancies – by better understanding the behavior of job seekers and job providers. Lines of reasoning proposed to explain the coexistence of unemployment and vacancies include the classical search models of McCall (1970), Mortensen (1970), and Lucas and Prescott (1974), as well as more recent models with workforce inertia due to Shimer (2007).[1] We complement this literature with a view that some degree of waiting is actually beneficial from a social welfare perspective as it enables market thickening — which in turn enables mismatch reduction. To illustrate this, consider the example of Shimer (2007), where some laid-off steel workers are not immediately given vacant positions as nurses. This may indeed be deemed optimal by a social planner when – by delaying their match – these nurse vacancies eventually are taken up by better nurses and the jobless steel workers find other jobs in the steel industry that might become available in the future.[2]

The second strand of related work comes from the matching literature and extends the canonical static matching framework to a dynamic setting.[3] As in the example of steel workers and nurses above, *mismatch* in dynamic environments may occur due to temporal inconsistencies, whereby, a posteriori, better matches were precluded by inferior matches that were formed earlier on. Therefore, some delay may be optimal from a social planner perspective to reduce mismatch. From a practical viewpoint, the challenge is to identify optimal mechanisms that thicken and clear the market in a way that balances these two objectives.

In this regard, Akbarpour et al. (2020), Ashlagi et al. (2023), Baccara et al. (2020), Loertscher et al. (2022), and Blanchet et al. (2022) break new ground in identifying optimal clearing schedules.[4] More precisely, Akbarpour et al. (2020), in the spirit of an organ exchange application such as the 'kidney exchange', identify the optimal mechanism to maximize the number of matches, that is, to minimize the number of agents perishing resulting from failing to get recipients matched with donors in time. In the model of Akbarpour et al. (2020), agents from both sides of the market arrive and leave stochastically and all carry identical match values, i.e., they are of the same type (in the spirit of each life being worth the same). However, any two agents are compatible, that is, can match, with some probability p.[5] The optimal mechanism identified minimizes the number of unmatched patients based on information concerning arrivals and departures, which may involve delaying compatible matches.[6] Without such information, greedy scheduling is always optimal in this setting. In fact, Ashlagi et al. (2023) show that greedy policies are generally optimal when each agent is either hard or easy to match, even if information about departure times is available when the kidney exchange market becomes large. Blanchet et al. (2022) study a related model where the utility from matching agents are general random variables. Focusing on departures they analyze optimal threshold policies. Kerimov et al. (2023) show in a general model, that if heterogeneity is limited to a bounded number of different types, greedy policies are approximately optimal. Recently, Bäumler et al. (2023) show a similar result for sparse compatibility graphs.

Baccara et al. (2020) and Loertscher et al. (2022) introduce to the latter models the notion of waiting times instead of perishing rates as in Akbarpour et al. (2020).[7] In Baccara et al. (2020), agents arrive in donor-recipient pairs and recipients are allowed to decline matches to remain in the market. This is motivated by the applications under scrutiny which include, among others, child adoption. As a result, one of the study's key focuses is on strategic incentives and their role in determining market outcomes. Their optimal clearing policy is discriminatory, in that it involves matching same-type pairs greedily, and delaying up to some threshold when there are only cross-type pairs in the market.

In theoretical computer science, the study of related questions dates back at least to the pioneering paper of Karp et al. (1990).[8] To the best of our knowledge, Emek et al. (2016) were the first in this strand of research to consider the scenario where all agents arrive on the market over time (instead of just one market side). They present a non-bipartite model where requests arrive stochastically from one of n different locations to study the performance of different algorithms in terms of worst-case matching and waiting cost.[9] In the setting of Emek et al. (2016), the specific match costs result from the distance between agents' locations so that, for a patient social planner, it is optimal to wait and only match agents who are at the same location. In another effort, the study of batching considers the optimal group size in scheduling tasks, where in general one side of the market arrives

[1] Note that Shimer (2007) terms his explanandum "mismatch" (as opposed to "waiting"), a term the matching literature uses to describe suboptimal matchings, which may be confusing.

[2] *Waiting* is explained behaviorally through inertia in Shimer (2007), that is, by the argument that steel workers stay close to their factories hoping that they reopen; Lucas and Prescott (1974) propose a different interpretation whereby waiting is due to the fact that steel workers must actively spend some time searching for these nursing jobs elsewhere.

[3] The canonical static frameworks underlying our analyses were pioneered by Koenig (1931), Egervary (1931), and Edmonds (1965); see also Gale and Shapley (1962) for matching with ordinal preferences.

[4] These are often inspired by some earlier papers on dynamic matching in organ exchange that focus on minimizing waiting time, absent of agents leaving (Zenios, 2002; Ünver, 2010). See also Bloch and Houy (2012), Kurino (2014), and Leshno (2022) who study related queuing models where one side of the market is already present (such as in the housing market).

[5] This can be modeled by means of a dynamically changing compatibility graph where edges represent feasible matches.

[6] Gurvich and Ward (2015) study a related queuing model where items arrive to different queues and a match needs to be made between a certain type of items.

[7] Su and Zenios (2004, 2005, 2006) study related one-sided queuing models with heterogeneous match values that are inspired by applications like the kidney exchange.

[8] Karp et al. (1990) and subsequent work – similar to its economic counterparts – focus on models with two market sides, where by contrast one side is typically present to begin with and incoming agents from the other side can only match with some of the present agents according to a compatibility graph (see Mehta (2013) for an overview and Aggarwal et al. (2011) for extensions to vertex-weighted matching).

[9] Azar et al. (2017) obtain additional results in terms of upper and lower bounds for the original model. Emek et al. (2019) obtain sharper results for a two-location model. There are also other extensions such as allowing for a stochastic graph (Anderson et al., 2015; Ashlagi et al., 2019).

while the other side is static (cf. Potts and Kovalyov (2000) and Pinedo (2012) for reviews). In contrast, we are concerned with the random arrival of demand and supply and matching event that only match one couple at a time.

Finally, motivated by ride-sharing applications, Ashlagi et al. (2017) extend the model of Emek et al. (2016) to a setting where two types of agents (clients and providers) independently arrive at n different locations. The cost of a matching is the sum of the distances plus the sum of the incurred waiting times. It is assumed that there is no information about the arriving agents and the distribution of arrivals across the n points. Using specific metrics (e.g., n equally spaced points on the unit interval) (Ashlagi et al., 2017) provide upper and lower bounds for the competitive ratio, that is, the ratio between their proposed randomized algorithms and the optimal solution. The bounds depend on the number of points n. By contrast, we study matching costs drawn from continuous distributions and the expected – as opposed to worst-case – performance, when the social planner knows the distribution. On the one hand, this allows us to identify optimal clearing schedules under different distributional assumptions and, on the other hand, requires us to introduce new tools that we believe can be useful in the future study of dynamic (matching) markets.

Contributions of the paper

Our paper examines dynamic markets with an infinite type space (in contrast to prior work that focuses on finite type spaces), a framework that we call the *dynamic clearing game*. In our setting, clients and providers arrive to the market stochastically and independently; the social planner has no information regarding the cost of matching couples currently in the market (but only the distribution from which they are drawn) and has no information about the future arrivals of individuals other their arrival rate. We posit that the cost of a matching event, that is, matching one (cheapest) couple is decreasing with market size. Then, the social planner must choose a clearing schedule that determines how long to wait between matching events. As such, the social planner is called to weigh, on the one hand, expected mismatches incurred from matching clients and providers suboptimally, and, on the other hand, the agents' waiting time.

We first establish a class of optimal clearing schedules for two extreme types of single-objective social planners – that is, for social planners who only care about minimizing waiting time (in which case greedy is best) or mismatch costs (resulting in unbounded delays), but not both at the same time. For the more general case of a social planner pursuing both objectives, we identify two regimes that are characterized as a function of the 'speed of improvement' of the cost of matching with respect to market size. One regime results in a *'quick or cheap'* dilemma where there is no 'free lunch,' that is, where no clearing schedule exists that is simultaneously approximately optimal along matching cost minimization and along waiting time minimization at the same time. Under this regime, we identify a unique breaking point where match costs starkly reduce while waiting times increase. The other regime permits a free lunch: a window of opportunity opens for clearing schedules where the objectives are simultaneously optimal.

We characterize how the optimality of clearing depends on the heterogeneity of matching costs. When heterogeneity is limited – i.e., if match costs come from a finite set –we show that greedy scheduling is approximately optimal. This is a result that is very much in line with results from related work discussed above (e.g., Ashlagi et al. (2023) and Kerimov et al. (2023)). By contrast, in the presence of infinitely many types, greedy clearing is generally not optimal. Jointly, these findings complement prior work showing that the quick-versus-cheap trade-off may indeed be more intricate. These results may then have consequences for applications that have been studied before (e.g., kidney exchange), where prior studies worked with binary types or limited heterogeneity and therefore identified greedy scheduling as optimal, in particular if other, richer match value metrics are used (e.g. potential years of life lost or disability-adjusted life years) that would result in wider spectrums of match costs.

To make the case of infinitely many types concrete, we study a micro-level model where match costs are distributed according to independent exponential random variables. Whilst the aforementioned general results hold for other distributions too, this model has the advantage of being tractable in closed form. Building on our no-free-lunch result, we fill the spectrum between matching cost and waiting time minimization and show how the speed at which the market is cleared determines the trade-off. We do so by introducing a class of clearing schedules covering a wide range of social planning desiderata between waiting time and matching cost, and achieving a continuous trade-off between the two. To explore the finer aspects of this trade-off, we introduce a utility model for the social planner whereby the associated utility of matching cost is of the same order as the agents' utility of waiting time. Under this model, we show that there exists a non-trivial clearing schedule achieving this balance, and we show that this schedule is effectively unique (up to asymptotic order considerations).

The key technical innovations of our paper concern the concurrent consideration of a continuum of types, independent arrivals, and incomplete information. In turn, these contributions rely on a range of tools from probability theory and disordered systems to obtain closed-form solutions. These underlying results are concerned with the expected matching cost for given instances of random, static assignment games. In particular, in static assignment games with the same number of clients and providers and $\exp(1)$ distributed edge weights, Aldous (2001) proved the long-standing conjecture that the expected minimum weight matching converges to $\pi^2/6$ as the number of players grows large. This result was later extended by Waestlund (2005) to assignment games with match costs drawn from non-identical exponential distributions.[10] By leveraging the techniques of Aldous (2001) and Waestlund (2005), we are able to compute the expected matching cost for every 'snapshot in time' of the dynamic clearing game. This provides strong foundations for our proofs which are then focused on estimating the fluctuations that result from the random arrival of clients and providers and their randomly drawn match costs. For the former it is critical for our results that we assume that clients and providers arrive with equal rates. To achieve this, we use several approximation techniques (in particular, the approximation of the arrival process by a continuous-time Wiener process), which allow us to port over several results from martingale limit theory (such as the law of the iterated logarithm).

Paper outline. The rest of the paper is structured as follows. In Section 2, we introduce the *dynamic clearing game*. Section 3 introduces the social planners' objectives, namely to avoid waiting time and mismatch, and provides preliminary results. Section 4 provides our first main result, namely that there are two regimes, one where free lunch is not achievable and one where it is achievable. Section 5 shows that for limited heterogeneity the greedy clearing schedule is approximately optimal. Section 6 is concerned with the analysis of the micro-level model where costs of matches are distributed according to exponential random variables. We analyze a natural selection of clearing schedules, which cover the whole range of possible trade-offs and show how to balance the two objectives. Finally, Section 7 concludes.

[10] To the best of our knowledge, the work of Walkup (1979) is the first to pose the question, while Mezard and Parisi (1987) conjectured the specific limit value. We also leverage the analyses of Buck et al. (2002) and Linusson and Waestlund (2004) who obtain results for the expected values of finite instances of the latter models, showing – as a byproduct – that the value is increasing with the number of agents. For a survey of this literature, we refer the reader to Krokhmal and Pardalos (2009).

2. The model

In this section, we introduce the model, which we shall refer to as the *dynamic clearing game*.

A dynamic two-sided market evolves in continuous time $\tau \in [0, \infty)$. At each tick of a Poisson clock with rate 1 an agent enters the market; this agent could be either a *client* or a *provider*, with equal probability. To keep track of the number of agents in both sides of the market, let $\mathcal{C}(\tau)$ and $\mathcal{P}(\tau)$ denote the set of clients and providers that have entered the market by time τ (and possibly already left again), and let $N_{\mathcal{C}}(\tau) = |\mathcal{C}(\tau)|$ and $N_{\mathcal{P}}(\tau) = |\mathcal{P}(\tau)|$ be the respective numbers thereof. Then, the number of agents that have entered the *short side of the market* will be written $N(\tau) = \min\{N_{\mathcal{C}}(\tau), N_{\mathcal{P}}(\tau)\}$.[11] Let S_τ be the difference of clients and providers who have arrived to the market until time τ, that is, $S_\tau = N_{\mathcal{C}}(\tau) - N_{\mathcal{P}}(\tau)$.

We consider a one-to-one matching market where each client is to be *matched* to at most one provider and vice versa; then, once a couple is matched, both agents leave the market. We focus on *matching events* (MEs) that match a single couple. We will write $A \equiv A(\tau)$ for the number of clients/providers that have been assigned a partner up to time τ, and $R(\tau) = N_{\mathcal{C}}(\tau) + N_{\mathcal{P}}(\tau) - 2A(\tau)$ for the number of unmatched agents at time τ. Further, write $M_{\mathcal{C}}(\tau) = N_{\mathcal{C}}(\tau) - A(\tau)$ for the number of clients in the market at time τ and $M_{\mathcal{P}}(\tau) = N_{\mathcal{P}}(\tau) - A(\tau)$ for the number of providers respectively.

Remark 1. Rather than modeling match costs directly by defining the distribution of each potential match cost, we take a macroscopic viewpoint and posit that the expected cost of matching a couple depends on the number of clients and providers currently in the market. This cost may – as we shall see in Sections 5 and 6 – be the expected cost of matching the cheapest couple. In practice, this cost can be learned from past data on clearing events.

Write the expected minimum cost as

$$g(M_{\mathcal{C}}, M_{\mathcal{P}}) \tag{1}$$

where $g : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is a non-increasing function (in either argument). For analytical convenience we define the function g on the real numbers – and shall use standard variables x, y when doing so, e.g., $g(x, y)$ or $g(x, x)$ –, but note that it is only the values on $\mathbb{N} \times \mathbb{N}$ which result in actually realized costs. Intuitively, g determines how the expected minimum cost of matching decreases as more players coexist in the market. For example, if $g(M_{\mathcal{C}}, M_{\mathcal{P}}) = \frac{1}{M_{\mathcal{C}} \cdot M_{\mathcal{P}}}$, yields the model studied in Section 6, where we assume exponentially distributed match costs.

Throughout the sequel, we assume the existence of a *social planner* who knows the expected cost of matching a couple dependent on how many players are in the game, $g(M_{\mathcal{C}}, M_{\mathcal{P}})$. They further observe the arrivals of agents to the market. The social planner has no other information. Due to this lack of information, the social planner has no basis to judge whether a particular agent arriving in the market is 'good' or 'bad', and is thus left with the challenge of choosing a clearing schedule with which to operate the market.

With all this in hand, a *clearing schedule* (CS) will be a rule that determines at which points in time $\tau \in (0, \infty)$ to trigger a matching event, possibly depending on $M_{\mathcal{C}}, M_{\mathcal{P}}$ (and $A(\tau)$).

We shall limit our analysis to *feasible* clearing schedules, that is, clearing schedules where the proportion of matched players approaches one in the longrun (more precisely, where $\lim_{\tau\to\infty} \frac{2A(\tau)}{N_{\mathcal{C}}(\tau)+N_{\mathcal{P}}(\tau)} = 1$).

[11] In a slight (but convenient) abuse of notation, we will sometimes write $N_{\mathcal{C}}(t)$, $N_{\mathcal{P}}(t)$, and $N(t)$ to denote respectively the number of clients, providers, and agents at the short side of the market when the tth agent enters the market – specifically, letting $\tau(t)$ denote the time at which the tth agent enters the market, we will write $N_{\mathcal{C}}(t) \equiv N_{\mathcal{C}}(\tau(t))$, etc.

Finally, we introduce the following, standard notation. For two functions, f and g, we will use the following asymptotic notations: $f(x) = \mathcal{O}(g(x))$ if $f(x) < c \cdot g(x)$ for some $c > 0$ constant and x sufficiently large. $f(x) = \Omega(g(x))$ is the inverse O notation ($f(x) > c \cdot g(x)$ for x sufficiently large). $f(x) = \Theta(g(x))$ if there exist two constants $k, K \geq 0$ and a positive integer x_0 such that $kg(x) \leq f(x) \leq Kg(x)$ for all $x \geq x_0$. For $g(x)$ non-zero $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$ and $f(x) = \omega(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$.

3. Objectives and preliminary results

The social planner aims to match clients and providers according to two (competing) considerations: (*a*) to reduce the coexistence of clients and providers (i.e., *waiting time*); and (*b*) to match clients and providers in a way that minimizes matching cost (i.e., *mismatch*). Beginning with the latter, define the *expected matching cost* for the first A couples as

$$\mathrm{cost}_{\mathtt{CS}}(A) \equiv \mathbb{E}\left[\sum_{k=1}^{A} g_{\mathtt{CS}_k}(M_{\mathcal{C}}, M_{\mathcal{P}})\right] \tag{2}$$

where $g_{\mathtt{CS}_k}(M_{\mathcal{C}}, M_{\mathcal{P}})$ is the minimum match cost of the kth matched couple under the clearing schedule $\mathtt{CS}$ and the expectation is taken with respect to the random arrival of clients and providers (recall that an arriving agent is either a client or a provider with equal probability). Similarly, define the *expected waiting time* of a clearing schedule until time T as

$$\mathrm{wait}_{\mathtt{CS}}(T) \equiv \mathbb{E}\left[\int_0^T R(\tau)\, d\tau\right] \tag{3}$$

where the expectation is taken with respect to the random arrival of agents in the market.

Our analysis begins with the case of a single-minded social planner. Specifically, we investigate which clearing schedule a social planner would employ if either only caring about the expected waiting time, or only caring about the expected matching cost.

First, a social planner who is optimizing the agents' expected waiting time will choose a clearing schedule which leaves no unmatched couples at any point in time. To do so, we will consider a 'greedy' clearing schedule, denoted $\mathtt{CS}_{\mathtt{greedy}}$, which performs a minimum weight matching whenever there is exactly one unmatched client/provider pair in the market. Second, a social planner who is optimizing the agents' expected matching cost will choose a clearing schedule which performs a single batch matching at time τ^* through consecutive matching events at that time. Then considering $\tau^* \to \infty$ yields the hypothetical 'patient' clearing schedule, denoted $\mathtt{CS}_{\mathtt{patient}}$ that will serve as a benchmark. This clearing schedule should be preferred by any social planner who is only concerned with the expected matching cost.

We shall say that a clearing schedule is *optimal with respect to matching cost minimization* if it minimizes matching cost among all feasible clearing schedules. Similarly, a clearing schedule is *optimal with respect to waiting time minimization* if it minimizes matching cost among all feasible clearing schedules.

Proposition 1. *The optimal clearing schedules for a single-objective social planner are:*

1. *The patient clearing schedule* $\mathtt{CS}_{\mathtt{patient}}$ *is optimal with respect to matching cost minimization; in particular, for all* $A \geq 1$*, write* $\mathrm{cost}_{\mathtt{patient}}(A)$.
2. *The greedy clearing schedule* $\mathtt{CS}_{\mathtt{greedy}}$ *is optimal with respect to waiting time minimization; in particular, for all* $\tau \geq 0$*, we have* $\mathrm{wait}_{\mathtt{greedy}}(\tau) = \Theta(\tau^{3/2})$.

In view of Proposition 1, the expected matching cost of $\mathtt{CS}_{\mathtt{patient}}$ and the expected waiting time of $\mathtt{CS}_{\mathtt{greedy}}$ will serve as the benchmark for comparing the matching cost and waiting time of any other clearing schedule.

Proof of Proposition 1. We prove our claims for each of the two clearing schedules separately.

Part 1: Matching cost minimization. The optimality of the patient clearing schedule with respect to matching cost minimization follows by observing that the matching resulting from any other clearing schedule can also be implemented under the patient clearing schedule. Therefore, the choice set of matchings of the patient clearing schedule is a superset of all choice sets of other clearing schedules. In particular this implies that the minimum cost matching attained by the patient clearing schedule is a lower bound for all possible matchings attained by other clearing schedules.

Part 2: Waiting time minimization. For our second assertion, note that, at any point in time, there are either no clients or no providers in the market. For all $T > 0$, we get:

$$\text{wait}_{\texttt{greedy}}(T) = \mathbb{E}\left[\int_0^T |S_\tau|\ d\tau\right] = \int_0^T \mathbb{E}\left[|S_\tau|\right]\ d\tau \tag{4}$$

where the latter equality holds by Tonelli (1909)'s theorem (since $|S_\tau|$ is non-negative).

Applying Tonelli (1909)'s theorem a second time, we can consider the case where the expectation with respect to the arrival times is taken first. To do so, consider an alternative (expected) process where at the fixed points in time $\tau = 1, 2, \ldots$ an agent arrives to the market and let $\bar{S}_\tau$ be the difference of clients and providers who have arrived to the market at time τ. Note that this process is the expectation of the process S_τ, where the expectation is taken with respect to the agents' random arrival times. We then have:

$$\mathbb{E}\left[\int_0^T |S_\tau|\ d\tau\right] = \int_0^T \mathbb{E}\left[|S_\tau|\right]\ d\tau = \int_0^T \mathbb{E}\left[|\bar{S}_\tau|\right]\ d\tau \tag{5}$$

It is known that for $\tau \to \infty$ the appropriately rescaled random walk $\bar{S}_\tau$ converges in distribution to the Wiener process W_τ (Kac, 1947). Thus, for large T, Eq. (5) gives

$$\mathbb{E}\left[\int_0^T |S_\tau| d\tau\right] = \Theta\left(\int_0^T \mathbb{E}\left[|W_\tau|\right]\ d\tau\right) = \Theta\left(\int_0^T \sqrt{\mathrm{Var}(W_\tau)}\, d\tau\right)$$
$$= \Theta\left(\int_0^T \sqrt{\tau}\, d\tau\right) = \Theta\left(\frac{2}{3}T^{3/2}\right).$$

The optimality of the greedy clearing schedule with respect to waiting time minimization follows by observing that the greedy clearing schedule minimizes the number of agents in the market at any given time. This is the case as there are never clients and providers in the market at the same time and the arrival process is independent of the clearing schedule. □

This concludes the analysis of a single-minded social planner who either only cares about the expected waiting time, or only the expected matching cost.

To go beyond the narrow view of a single-minded social planner, we define the expected matching cost ratio and the expected waiting time ratio:

1. The *expected matching cost ratio* of a clearing schedule CS is

$$\alpha \equiv \alpha(A) = \frac{\text{cost}_{\texttt{CS}}(A)}{\text{cost}_{\texttt{patient}}(A)}. \tag{6}$$

2. The *expected waiting time ratio* of a clearing schedule CS is

$$\beta \equiv \beta(\tau) = \frac{\text{wait}_{\texttt{CS}}(\tau)}{\text{wait}_{\texttt{greedy}}(\tau)}. \tag{7}$$

We restrict our attention to clearing schedules that can be characterized by a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that the kth ($k \in \mathbb{N}$) couple is matched when $\lceil f(k) \rceil$ agents are on the short side of the market (for example, the greedy clearing schedule is described by $f \equiv 1$). In particular, note that this rules out non-deterministic clearing schedules and clearing schedules that depend on the exact (and not only expected) size of the long side of the market. Denote a clearing schedule that is defined via a function f by $\texttt{CS}_f$. Given that we want to compare the asymptotic behavior of different clearing schedules, we further restrict our analysis on a natural class of functions introduced by Hardy (1910) which make such comparisons possible. Specifically, each function in this class is defined, for all $x \geq 0$, by a finite combination of the basic arithmetic operations (addition, multiplication, raising to a power, and their inverses), operating on the variable x and on real constants. Hardy (1910, Theorem, page 18) shows that for any two such functions, f and g, either $f = \omega(g)$, $f = \Theta(g)$, or $f = o(g)$.

Table 1
Overview of the various clearing schedules considered in the sequel. Note that all schedules other than $\texttt{CS}_{\texttt{FCFS}}$ match the couple with the minimum match cost at each matching event.

Schedule	Description
$\texttt{CS}_{\texttt{FCFS}}$	Match players as soon as possible a la first-come, first-served (FCFS)
$\texttt{CS}_{\texttt{greedy}}$	Match players as soon as possible (as best as possible, not FCFS)
$\texttt{CS}_\gamma$	Match the kth couple when $\Theta(k^\gamma)$ players are on the short side of the market ($0 \leq \gamma \leq 1$); γ determines the 'rate' of matching
$\texttt{CS}_{\texttt{patient}}$	Match players optimally after everyone has arrived
$\texttt{CS}_{\texttt{balanced}}$	Match the kth couple when $\Theta(k^{1/2}(\log k)^{1/3})$ players are on the short side of the market

Interpolating between the two 'extreme' clearing schedules – the patient clearing schedule $\texttt{CS}_{\texttt{patient}}$ (which minimizes mismatches) and the expected waiting time of the greedy clearing schedule $\texttt{CS}_{\texttt{greedy}}$ (which minimizes waiting times) –, we will also consider a class of clearing schedules where the social planner waits for some length of time to accrue some intermediate number of agents on both sides of the market. Concretely, we shall study clearing schedules that match the kth couple when $N - k = f(k)$, i.e., when $f(k)$ players are on the short side of the market.[12]

For concreteness, we restrict ourselves to clearing schedules of the form

$$f(k) = \Theta(k^\gamma) \quad \text{for some } \gamma \in [0, 1]. \tag{8}$$

For $\gamma = 0$, the induced clearing schedules match players once a constant threshold is reached; in particular, the greedy schedule is recovered when $f \equiv 1$ (corresponding to $\gamma = 0$). More generally, we shall denote clearing schedules of the above form by $\texttt{CS}_\gamma$ and write $\texttt{CS}_{1/2}$ for the clearing schedule with $\gamma = 1/2$. Similarly we shall use the notation α_γ for the expected matching cost ratio of $\texttt{CS}_\gamma$ and β_γ for the expected waiting time ratio of $\texttt{CS}_\gamma$. Table 1 summarizes all clearing schedules analyzed in this vein (including a 'balanced' schedule, $\texttt{CS}_{\texttt{balanced}}$, that we discuss in Section 6.2) and another natural schedule based on the principle of *first-come, first-served* (FCFS), i.e., when agents are matched as soon as possible on a first-come, first-served basis; denote this schedule by $\texttt{CS}_{\texttt{FCFS}}$.[13]

Finally, we say that a clearing schedule CS has *finite expected matching cost ratio* if $\limsup_{A(\tau)>0} \alpha(A(\tau)) < \infty$ and has *critical rate matching cost ratio* if $\limsup_{A(\tau)>0} \alpha(A(\tau)) = \Theta(\log(A))$; we say that a clearing schedule has *finite expected waiting time ratio* if $\limsup_{\tau>0} \beta(\tau) < \infty$. We say that a clearing schedule achieves *free lunch* if it has both finite expected matching cost ratio and finite expected waiting time ratio.

[12] Recall that $N = \min\{N_C, N_P\}$.

[13] Note that $\texttt{CS}_{\texttt{FCFS}}$ differs from $\texttt{CS}_{\texttt{greedy}}$ in terms of who is matched with whom (first-come, first-served vs. minimum cost matching) but not regarding when a matching event occurs. As such, given that $\texttt{CS}_{\texttt{FCFS}}$ does not take into account matching costs, it is not reasonable to expect that it will perform well on any dimension other than the agents' expected waiting times. On the other hand, it exhibits 'fairness' relative to the agents' arrival times, a feature which is crucial in many applications. Indeed, this may be a desirable feature in applications such as processor time requests in distributed computing. We shall leave extensions of our analyses to include fairness considerations for future work.

4. Two regimes: Free lunch versus no free lunch

The first question that arises is whether there exists a clearing schedule that is optimal along *both* dimensions (at least, asymptotically), that is, whether there exists a free lunch or not. The answer to this question is nuanced: it depends on the 'speed of the improvement' of the match cost with respect to the size of the market. The asymptotic behavior of the market will be captured by the rate at which the expected minimum matching cost $g(M_{\mathcal{C}}, M_{\mathcal{P}})$ vanishes as a function of $M_{\mathcal{C}}, M_{\mathcal{P}} \to \infty$. Theorem 1 below makes this intuition precise and identifies a specific threshold beyond which it *is* possible to get a free lunch.

Theorem 1. *Suppose that for $M_{\mathcal{C}} = \Theta(x), M_{\mathcal{P}} = \Theta(x)$ the expected minimum matching cost decays as $g(M_{\mathcal{C}}, M_{\mathcal{P}}) = \Theta(1/x^{\delta})$ for some $\delta > 1$. Then:*

(i) For $1 < \delta \leq 2$ there is no free lunch. In particular, a critical rate clearing schedule, that is, the clearing schedule with expected matching cost ratio $\Theta(\log(A))$, is given by $\mathtt{CS}_{1/\delta}$.
(ii) For $\delta > 2$, free lunch exists. In particular, the clearing schedules $\mathtt{CS}_{\gamma}$ with $\gamma \in (\frac{1}{\delta}, \frac{1}{2}]$ guarantee that the expected matching cost and waiting time ratios are both finite.

Theorem 1 finds two regimes, one where a free lunch is not possible and one where it is. Importantly, it shows that the crucial element is the speed of the improvement of the match cost as the market size grows. This is because for quickly decaying matching costs it becomes easier to choose a 'good' schedule, and thus the 'window of opportunity' is increasing in the derivative of g. One can build intuition for this result by reasoning about market settings that differ in terms of match cost variability: in markets where 'good' matches are 'rare', thickening the market by waiting will only lead to a meaningful positive effect in terms of expected match cost reduction when waiting for a long time. By contrast, when 'good' matches are 'common', match costs reduce in expectation with much less delay, thus making it more likely for a mechanism designer to achieve a free lunch. Importantly, the clearing schedule $\mathtt{CS}_{\mathtt{greedy}}$ is never optimal when dealing with fully heterogeneous types distributions (i.e., continuous), independent of the match cost distribution at hand (see Section 5, for an analysis of limited heterogeneity).

Note that we restrict our analysis to $\delta > 1$, guaranteeing that the patient clearing schedule has finite expected matching cost, i.e., $\mathrm{cost}_{\mathtt{patient}} < \infty$. This is the case, as for the patient clearing schedule the expected matching cost ratio is then bounded by $\zeta(\delta)$, where ζ is the Riemann zeta function. We can thus focus our attention on the numerator of the expected matching cost ratio.

Let $t(k, f(k))$ be the stopping time for the event that for the kth time at least $f(k)$ clients and $f(k)$ providers are in the market, assuming that every time this is the case one client and one provider are removed.

Proof of Theorem 1. (i) We first consider the upper bound. Given $g(\Theta(x), \Theta(x)) = \Theta\left(\frac{1}{x^{\delta}}\right)$ and since g is increasing in both arguments we have (where the conditioning determines the clearing schedule):

$$\mathbb{E}\left[\sum_{k=1}^{A} g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor\right] \leq \sum_{k=1}^{A} g(\lfloor k^{1/\delta} \rfloor, \lfloor k^{1/\delta} \rfloor) = \Theta\left(\sum_{k=1}^{A} \frac{1}{k}\right) = \Theta(\log A) \tag{9}$$

where the last inequality follows from the bounds for the harmonic series. Thus, given the optimal clearing schedule has finite matching cost, the expected matching cost ratio is smaller than $O(\log A + 1)$.

For the lower bound note that $\mathbb{E}[t(k, k^{1/\delta})] < 10k$ by Appendix A, Lemma 4 and by recalling that $\delta > 1$. Thus, combining Jensen (1906)'s inequality ($\frac{1}{x}$ is convex) with Markov's inequality, the variation bounds from Appendix B.2, Lemma 2, and recalling that $\delta \leq 2$ we get:

$$\begin{aligned}
&\sum_{k=1}^{A} \mathbb{E}\left[g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor\right] \\
&= \sum_{k=1}^{A} \mathbb{E}\left[g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor,\ t(k, k^{1/\delta}) < 20k\right] \cdot \mathbb{P}\left[t(k, k^{1/\delta}) < 20k\right] \\
&+ \sum_{k=1}^{A} \mathbb{E}\left[g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor,\ t(k, k^{1/\delta}) > 20k\right] \cdot \mathbb{P}[t(k, k^{1/\delta}) > 20k] \\
&\geq \sum_{k=1}^{A} \mathbb{E}\left[g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor,\ t(k, k^{1/\delta}) < 20k\right] \cdot \frac{1}{2} \\
&= \Theta\left(\sum_{k=1}^{A} g(\lfloor k^{1/\delta} \rfloor, \lfloor k^{1/\delta} \rfloor)\right) = \Theta\left(\sum_{k=1}^{A} \frac{1}{k}\right) = \Theta(\log A)
\end{aligned} \tag{10}$$

where we used that $M_{\mathcal{C}} = \Theta(M_{\mathcal{P}})$ given $|S_t| = |N_{\mathcal{C}}(t) - N_{\mathcal{P}}(t)| = |M_{\mathcal{C}}(t) - M_{\mathcal{P}}(t)| = O(\sqrt{20k})$ since we are in the case $t(k, k^{1/\delta}) < 20k$ and by the assumption that for all λ, μ we have $g(\lambda \cdot x, \mu \cdot x) = \Theta(x^{\delta})$. Thus, given the optimal clearing schedule has finite matching cost, the expected matching cost ratio is $\Omega(\log A)$, concluding the proof together with the upper bound.

To summarize, for $\delta \in (1, 2]$ a critical rate clearing schedule is given by $\mathtt{CS}_{1/\delta}$. Thus – up to $O(1)$ differences – the critical rate clearing schedules are given by $\mathtt{CS}_{\gamma}$ with $\gamma \in (0, 1/2]$ and by Appendix B.3, Theorem 5, the expected waiting time ratio for these schedules is not finite. We conclude that there is no free lunch.

(ii) By Appendix B.3, Theorem 5 the expected waiting time ratio is finite for all clearing schedules $\mathtt{CS}_{\gamma}$ with $\gamma \leq \frac{1}{2}$.

We upper bound the expected matching cost ratio for the clearing schedule $\mathtt{CS}_{\gamma}$ for $\gamma > \frac{1}{\delta}$: For the upper bound we have with $g(\Theta(x), \Theta(x)) = \Theta\left(\frac{1}{x^{\delta}}\right)$:

$$\mathbb{E}\left[\sum_{k=1}^{A} g(M_{\mathcal{C}}, M_{\mathcal{P}}) \middle| \min\{M_{\mathcal{C}}, M_{\mathcal{P}}\} = \lfloor k^{1/\delta} \rfloor\right] \leq \sum_{k=1}^{A} g(\lfloor k^{1/\delta} \rfloor, \lfloor k^{1/\delta} \rfloor) = \Theta\left(\sum_{k=1}^{A} \frac{1}{k^{\gamma \cdot \delta}}\right) = \Theta(1) \tag{11}$$

where the last identity holds since $\gamma > \frac{1}{\delta}$.

Thus, for given δ the clearing schedules $\mathtt{CS}_{\gamma}$ with $\gamma \in (\frac{1}{\delta}, \frac{1}{2}]$ guarantee that the expected matching cost ratio and waiting time ratio are both finite, i.e., free lunch. □

5. Optimality of greedy policies for limited heterogeneity

As discussed in Section 1 prior work has identified that greedy clearing schedules are approximately optimal (e.g., Ashlagi et al. (2023) and Kerimov et al. (2023)). Why then do we find that greedy clearing schedules are never optimal in our model when the social planner attaches some utility to reducing waiting time? The answer lies in the assumption about the distribution of types. Earlier work either assumes that agents are either compatible or incompatible, and/or that their match costs are drawn from a finite set, that is, they exhibit limited heterogeneity. Let us consider match costs that are drawn from a finite set with minimum match cost $\underline{w} > 0$ and maximum match cost $\overline{w} \geq \underline{w}$ and the probability of a match having cost $\underline{w}$ is $p > 0$.

The following theorem gives – in our modeling context – an analogous for the above mentioned results.[14]

[14] Note that the expected matching cost, $\mathrm{cost}_{\mathtt{patient}}(A)$, is unbounded. However, we can still consider the limit of the expected matching cost ratio.

Theorem 2. *When there is a finite set of positive match costs,* $\mathtt{CS}_{\mathtt{greedy}}$ *provides a free lunch.*

Intuitively, greedy policies are optimal, when the gains from waiting are limited. Moreover, as is the case in our model due to random arrivals, if one of the market sides is growing with time, the probability that a least-cost matching is available for a new entrant in the empty market side quickly tends to one, i.e., that the agent can be matched with someone at cost $\underline{w}$. This renders $\mathtt{CS}_{\mathtt{greedy}}$ (approximately) optimal in terms of matching costs.

Proof. For the greedy clearing schedule, the kth match happens when the minimum of the number of clients and providers who already arrived to the market is k, that is, at time $t(k,1)$. The expected weight of the kth match depends on the number of players currently present on the long side of the market (since on the short side there is only one player). Let this random variable be denoted by $|S_{t(k,1)}| + 1$ (see Appendix B.2, Definition 1 for a formal definition). Further, note that $\mathrm{cost}_{\mathtt{patient}}(A) \geq A \cdot \underline{w}$.

The expected matching cost ratio is thus upper bounded by:

$$\frac{1}{A\underline{w}}\mathbb{E}[\sum_{k=1}^{A}\underbrace{(1-(1-p)^{|S_{t(k,1)}|+1})\underline{w}}_{<\underline{w}}+(1-p)^{|S_{t(k,1)}|+1}(\overline{w}-\underline{w})]. \tag{12}$$

Given that A is fixed, the latter is upper-bounded by:

$$\frac{1}{A\underline{w}}\sum_{k=1}^{A}\left[\underline{w}+\mathbb{E}[(1-p)^{|S_{t(k,1)}|+1}]\cdot(\overline{w}-\underline{w})\right]$$
$$=1+\frac{(\overline{w}-\underline{w})}{A\underline{w}}\cdot\sum_{k=1}^{A}\mathbb{E}[(1-p)^{|S_{t(k,1)}|+1}] \tag{13}$$

Since the function $(1-p)^x$ is convex, Jensen (1906)'s inequality implies that the latter is upper-bounded by:

$$1+\frac{(\overline{w}-\underline{w})}{A\underline{w}}\cdot\sum_{k=1}^{A}(1-p)^{\mathbb{E}[|S_{t(k,1)}|+1]} \tag{14}$$

Next, by Lemma 2 we have $\mathbb{E}[|S_{t(k,1)}|] \geq \frac{2\pi}{e^2}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{k} > 0.5\cdot\sqrt{k}$. Thus Eq. (14) is upper-bounded by:

$$1+\frac{(\overline{w}-\underline{w})}{A\underline{w}}\cdot\sum_{k=1}^{A}(1-p)^{0.5\cdot\sqrt{k}} \tag{15}$$

By the ratio test the latter sum in Eq. (15) converges to a finite number (as A goes to infinity). This concludes that the expected matching cost ratio converges to one. Finally, it immediately follows that there exists a free lunch, as $\mathtt{CS}_{\mathtt{greedy}}$ is optimal with respect to waiting time by Proposition 1. □

6. Exponentially distributed match weights

Section 4 identified when free lunch exists dependent on the speed of improvement in match costs as the market grows. Further, Section 5 identified why prior work had found greedy clearing schedules as optimal, namely due to a finite set of possible match costs. We next consider a specific match cost distribution that arises from exponentially distributed match costs. Moreover, it turns out that it is exactly at the boundary of the free lunch versus no-free lunch regimes identified in Theorem 1 (that is, $\delta = 2$).

Suppose that the quality of a (candidate) pair is characterized by an inherent, heterogenous *match parameter* λ_{ij} where a higher parameter will represent a lower expected match cost. Match costs are independently and exponentially distributed with rate λ_{ij}. Specifically, we posit that the *match cost* $w_{ij} > 0$ when client $i \in \mathcal{C}(\tau)$ is matched to provider $j \in \mathcal{P}(\tau)$ is an independent draw from an exponential distribution of rate λ_{ij} for any time τ, that is, $w_{ij} \sim \exp(\lambda_{ij})$. For example, a popular model assumes that λ_{ij} is composed by additively separable components describing the agents' types and a couple-specific term depending possibly on both the identity of the agents and their types (Kanoria et al., 2018). For generality, our only assumption regarding the rate parameters λ_{ij} is that they are bounded from below by $\underline{\lambda}$, from above by $\overline{\lambda}$, and have mean value λ, that is, $\lambda = \lim_{\tau\to\infty}\left[N_{\mathcal{C}}(\tau)\cdot N_{\mathcal{P}}(\tau)\right]^{-1}\sum_{i=1}^{N_{\mathcal{C}}(\tau)}\sum_{j=1}^{N_{\mathcal{P}}(\tau)}\lambda_{ij}$ almost surely. In particular, parameters do not need to be independent, but we shall assume that they are fixed at time 0.

Remark. As mentioned by Aldous (2001) and developed in detail by Janson (1999, Section 2) generalizations to larger classes of distributions are easily obtained. In particular, the condition only requires the distribution function, write F, to be 'well behaved' at the left limit, namely $F(t)/t \to 0$ for $t \searrow 0$. Note that more generally, our results rely on the assumption that the expected matching cost is decreasing in the size of the market.

We assume that the social planner knows that match costs are drawn according to above defined exponential distributions, but does not observe their realizations. Thus, a clearing schedule will again be a rule that determines at which points in time $\tau \in (0,\infty)$ to trigger a matching event, possibly depending on $M_{\mathcal{C}}, M_{\mathcal{P}}$ (and $A(\tau)$).

We first observe that Theorem 1 applies to the model with exponentially distributed match costs:

Corollary 1 (*No Free Lunch*)**.** *With exponentially distributed match costs, there exists no clearing schedule simultaneously achieving finite ratios for both expected matching cost and waiting time. In particular,* $\delta = 2$.

Proof of Corollary 1. By Appendix B.1, Lemma 5 the patient clearing schedule has finite expected matching cost, therefore Theorem 1 applies. Next, the expected minimum of $M_{\mathcal{C}} \cdot M_{\mathcal{P}}$ independent $\exp(1)$ random variables is equal to $\frac{1}{M_{\mathcal{C}}\cdot M_{\mathcal{P}}}$ (see Appendix B.2, Lemma 1). Thus $g(M_{\mathcal{C}}, M_{\mathcal{P}}) = \frac{1}{M_{\mathcal{C}}\cdot M_{\mathcal{P}}}$ and if $M_{\mathcal{C}} = \Theta(x)$ and $M_{\mathcal{P}} = \Theta(x)$ we have $g(M_{\mathcal{C}}, M_{\mathcal{P}}) = \Theta(1/x^2)$. Hence $\delta = 2$. □

6.1. Interpolating between waiting time and matching cost

With the no free lunch result at hand, we here show how the trade-off evolves with clearing schedules parametrized by the speed at which the market is cleared. This allows to choose a clearing schedule according to a social planning desiderata from a broad spectrum interpolating between matching cost and waiting time. Our results (in terms of each schedule's expected matching cost and waiting time ratio) are summarized in Table 2: as can be seen, the family of schedules under study captures the full range between schedules that are 'good' relative to mismatches and 'bad' relative to waiting times, and vice versa. The formal statements of the results presented in Table 2 and their proofs are relegated to Appendices B.2 and B.3 in order to streamline our presentation.

In view of these results, the clearing schedule $\mathtt{CS}_{1/2}$ can be seen as a *phase transition* between two markedly different regimes. On the one hand, for $\gamma < 1/2$, the expected matching cost ratio $\alpha(A)$ grows as a power law in A while the expected waiting time ratio $\beta(\tau)$ is finite. On the other hand, for $\gamma > 1/2$, we have a finite expected matching cost ratio but an expected waiting time ratio that grows polynomially. Finally, at the critical point $\gamma = 1/2$, the expected matching cost ratio grows to infinity for large A, but at a slow, logarithmic rate ($\Theta(\log A)$). Notably, the phase transition at $\gamma = 1/2$ signifies a discontinuity of the expected matching cost ratio, so it is a *first-order* phase transition; by contrast, the expected waiting time ratio exhibits no such discontinuity, signifying a *second-order* phase transition.

The infinite matching cost ratio vis-a-vis the finite waiting time ratio for $\gamma = 1/2$ suggests that further fine-tuning should be possible and, indeed, the 'balanced' schedule $\mathtt{CS}_{\mathtt{balanced}}$ (which we define and discuss in Section 6.2) reduces the growth of the expected matching cost ratio

Table 2
The range of expected matching cost and waiting time ratios; $\mathtt{CS}_{\mathtt{balanced}}$ is discussed in Section 6.2. Recall from Eqs. (6) and (7) the expected matching cost ratio $\alpha \equiv \alpha(A) = \frac{\mathrm{cost}_{\mathtt{CS}}(A)}{\mathrm{cost}_{\mathtt{patient}}(A)}$ ($\mathrm{cost}_{\mathtt{patient}}(A) = \Theta(1)$) and the expected waiting time ratio $\beta \equiv \beta(\tau) = \frac{\mathrm{wait}_{\mathtt{CS}}(\tau)}{\mathrm{wait}_{\mathtt{greedy}}(\tau)}$ ($\mathrm{wait}_{\mathtt{greedy}}(\tau) = \Theta(\tau^{3/2})$).

Schedule	Rate of matching	Matching cost ratio, α	Waiting time ratio, β
$\mathtt{CS}_{\mathtt{FCFS}}$	FCFS	$\Theta(A)$	1
$\mathtt{CS}_{\mathtt{greedy}}$	Greedy	$\Omega(A^{1/2})$ ↑	1 ↓
$\mathtt{CS}_{0\le\gamma<1/2}$	Subcritical	$\Omega(A^{1/2-\gamma})$ ↑	$\Theta(1)$ ↓
$\mathtt{CS}_{1/2}$	Critical	$\Theta(\log A)$ ↑	$\Theta(1)$ ↓
$\mathtt{CS}_{1/2<\gamma\le 1}$	Supercritical	$\Theta(1)$ ↑	$\Theta(\tau^{\gamma-1/2})$ ↓
$\mathtt{CS}_{\mathtt{patient}}$	Patient	1 ↑	$\Theta(\tau^{1/2})$ ↓
$\mathtt{CS}_{\mathtt{balanced}}$	Balanced matching	$\Theta((\log A)^{1/3})$	$\Theta((\log A)^{1/3})$[a]

[a] For technical reasons this result is stated in terms of the number of matched couples, cf. Remark 2.

by a factor of $(\log A)^{2/3}$ while increasing the expected waiting time ratio $\beta(\tau)$ by a factor of $(\log \tau)^{1/3}$. In a sense (that we shall make precise in the next section) this is as close as we can get to a free lunch in this setting.

6.2. A balanced social planner

Until now, we have analyzed clearing schedules based on the trade-off between waiting time and matching cost, but without explicitly comparing the two. In this section, we shall commit to a specific class of utility functions to make an explicit comparison between these otherwise incomparable quantities.

To that end, let $u(\cdot)$ denote the expected utility (or 'welfare') of the social planner given a specific clearing schedules. Assume further that the functions expressing this utility depend on both the expected matching cost and the expected waiting time via the additively separable expression

$$u(\mathtt{CS}) = u_{\mathrm{cost}}(\alpha_{\mathtt{CS}}) + u_{\mathrm{wait}}(\beta_{\mathtt{CS}}) \tag{16}$$

To make comparisons between the utility components u_{cost} and u_{wait} we shall first consider their respective maximum values. It is then natural to assume that u_{cost} is maximal for the patient clearing schedule (which minimizes matching cost) and that u_{wait} is maximal for the greedy clearing schedule (which minimizes waiting time). We shall thus assume that the two maxima are of the same order, viz.,

$$\sigma \cdot u_{\mathrm{cost}}(\alpha_{\mathtt{patient}}) = (1-\sigma) \cdot u_{\mathrm{wait}}(\beta_{\mathtt{greedy}}) \tag{17}$$

where $\sigma \in (0,1)$ is a constant factor that specifies the relative importance of the disutilities from mismatching versus waiting. Naturally, we require that the social planner seeks to minimize both the costs of matching and the agents' waiting time. As such, we make the assumption that u_{cost} is a concave function that decreases in the expected matching cost, and u_{wait} is a concave function that decreases in the expected waiting time.

In view of all this, a social planner maximizes their utility when the disutilities from mismatching and waiting display similar growths for large τ; we then say that the social planner is *balanced*. That is, for a given clearing schedule $\mathtt{CS}$ with expected matching cost ratio α and expected waiting time ratio β, it holds that

$$u_{\mathrm{cost}}(\alpha_{\mathtt{CS}}) = \Theta(u_{\mathrm{wait}}(\beta_{\mathtt{CS}})) \quad \text{whenever} \quad \alpha_{\mathtt{CS}} = \Theta(\beta_{\mathtt{CS}}) \tag{18}$$

In this general context, we obtain the following result governing balanced social planning:

Theorem 3. *Let $\mathtt{CS}_{\mathtt{balanced}}$ be the clearing schedule that matches the kth couple when $N-(k-1) = \lceil k^{1/2}(\log(k+1))^{1/3}\rceil$ players are on the short side of the market. $\mathtt{CS}_{\mathtt{balanced}}$ is optimal, in particular the expected matching cost and waiting time ratios incurred by $\mathtt{CS}_{\mathtt{balanced}}$ are both $\Theta((\log A)^{1/3})$. Moreover, a clearing schedule $\mathtt{CS}_f$ is optimal if and only if $f(k) = \Theta(k^{1/2}(\log(k+1))^{1/3})$.*

The proof is provided in Appendix B.4. It relies on studying the difference of clients and providers as a random walk and bounding stopping times until matching events occur. To do so we require the use of the law of iterated logarithm (Khintchine, 1924) and a number of algebraic manipulations.

Remark 2. For technical reasons we state our result in terms of the number of matched couples ($A = A(\tau)$). Note that, for any clearing schedule where the proportion of matched players increases over time (more precisely, where $\lim_{\tau\to\infty} \frac{2A(\tau)}{N_C(\tau)+N_P(\tau)} = 1$), A is growing at the same rate as τ.

The clearing schedule $\mathtt{CS}_{\mathtt{balanced}}$ achieves – up to constants – an exact balance between the matching cost ratio and the waiting cost ratio, thus closing the gap in Table 2. Note that, up to logarithmic factors, the balanced clearing schedule is close to the clearing schedule $\mathtt{CS}_{1/2}$ which signifies a first-order phase transition for the expected matching cost ratio. As discussed earlier, $\mathtt{CS}_{1/2}$ only signifies a second-order phase transition for the expected waiting time ratio, thus explaining the gap between $\mathtt{CS}_{\mathtt{balanced}}$ and $\mathtt{CS}_{1/2}$. In practice however, $\mathtt{CS}_{1/2}$ seems to be a reasonable approximation for a balanced social planner.

7. Discussion

In this paper, we studied the *dynamic clearing game*, where heterogeneous clients and providers arrive at random points in time to be matched. We studied the trade-off a social planner is facing between two competing objectives: (a) to reduce players' *waiting time* before getting matched; and (b) to form efficient pairs to reduce *matching cost*.

We initially abstract away from modeling match costs directly and take a macroscopic viewpoint. Positing that the cost of matching a couple depends on the number of clients and providers currently in the market we identify two regimes. One, where no free lunch holds, the other, where there is a window of opportunity to be optimal along both dimensions, that is free lunch. When there is no free lunch, we identified a unique breaking point where a stark reduction in matching cost compared to a stark increase in waiting time occurs. For limited heterogeneity we find that greedy clearing schedules are optimal; however for full heterogeneity they never are.

We then studied a micro-founded model for match costs, namely exponentially distributed match costs. We showed how varying clearing schedules resolve the trade-off between waiting time and matching cost. In line with recent works by Ashlagi et al. (2017, 2019) and many others, we focused on a concrete class of social welfare functions that weigh costs from waiting versus matching on a comparable scale and identify the optimal clearing schedule, namely, the clearing schedule that matches the kth couple when $\Theta(\sqrt{k}(\log k)^{1/3})$ players are on the short side of the market.

There are multiple directions in which our analysis could be extended. Perhaps the most evident avenue for future research is to model market participation behavior game-theoretically, which would lead to new strategic considerations and probably induce other matchings (see, e.g., Baccara et al. (2020)). This analysis could be pursued in more applied contexts, for instance relating to our motivating example of a labor market with a central employment bureau, where waiting times could be interpreted as benefits payable by the bureau. An unemployed worker might forgo some of these benefits by (repeatedly) rejecting matches. This is the case because longer waiting, even though borne out of strategic behavior, may improve the match quality (reducing matching cost).

A second route for further investigation is to enlarge the options of the social planner in terms of clearing schedules. For one, the social planner could be learning from market observations about the distribution of match costs, which incidentally we may also allow to follow other, more general classes of distributions. This would allow the social planner to formulate more sophisticated clearing schedules that incorporate match costs between players that are currently in the market. In particular, if the social planner learns that a given agent may be 'hard to match', then it might be sensible to match that agent directly and not incur further waiting cost. Furthermore, the social planner may want to match more than one couple at a time.

A third route for further inquiry is to study different arrival (and departure) processes. For instance, it would be interesting to consider an urn model with delayed replacement or a birth–death process where clients and providers leave the market at some point if not being matched.

The study of dynamic market institutions is clearly fascinating, with tremendous scope for progress in (old and new) applications, where research has only just started. Our contribution has been to go beyond binary match values, and to identify breaking points in this general framework. We hope that our framework is able to provide fertile ground for further research, both theoretical and applied to real-world market contexts, in particular as regards thinking about whether the kinds of breaking points we describe are relevant in the optimal design of such markets.

CRediT authorship contribution statement

Panayotis Mertikopoulos: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Heinrich H. Nax:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Bary S.R. Pradelski:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

We hereby confirm that over the past three years we have not had any relevant conflicts of interest, such as employment, consulting, stock ownership.

Data availability

No data was used for the research described in the article.

Acknowledgments

The paper has benefited from helpful comments by Vahideh Manshadi, Igal Milchtaich, Jonathan Newton, Sven Seuken, Philipp Strack, and from anonymous referees. We thank all of them, as well as seminar participants at Bar-Ilan University, the Paris Game Theory Seminar, the INFORMS Workshop on Market Design 2019 and the 21st ACM Conference on Economics and Computation (EC'20). We also thank Simon Jantschgi and Dimitrios Moustakas for careful proof-reading. All remaining errors are our own. PM benefited from the support of the COST Action CA16228 "European Network for Game Theory" (GAMENET). HN benefited from the support of the ERC Advanced Investigator Grant 'Momentum' (No. 324247) and from an SNSF Eccellenza grant ('Markets and Norms'). BP benefited from the support of the ANR grants 15-IDEX-02, 11LABX0025-01, ALIAS and the Oxford-Man Institute of Quantitative Finance.

Appendix A. Technical lemmas

Definition 1. Let $X_1, X_2, \dots$ be iid random variables with $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$.

- Let $S_k = \sum_{i=1}^{k} X_i$.
- Let $t(k, C)$ be the stopping time for the event that for the kth time at least C clients and C providers are in the market, assuming that every time th latter is the case one client and one provider are removed.
- Let $S_{t(k,1)} = \sum_{i=1}^{t(k,1)} X_i$.

Lemma 1. *Let $w_{ij} \sim \exp(\lambda_j)$ for $j = 1, 2, \dots, N$, be a family of independent exponentially distributed random variables. Then*

$$\min\{w_{i1}, w_{i2}, \dots, w_{iN}\} \sim \exp\left(\sum_{j=1}^{N} \lambda_j\right). \tag{19}$$

In particular, if for all j, $\lambda_j = 1$, then $\mathbb{E}[\min_j w_{ij}] = \frac{1}{N}$.

Proof. This proof is standard but we repeat it for the sake of completeness. The random variable w_{ij} has cumulative distribution function

$$F_{w_{ij}} = \mathbb{P}\left(w_{ij} \le x\right) = 1 - e^{-\lambda_j x} \quad \text{for all } x > 0 \text{ and all } j = 1, 2, \dots, N. \tag{20}$$

Now, define the random variable $Y = \min\{w_{i1}, w_{i2}, \dots, w_{iN}\}$. Then, the cumulative distribution function of Y is

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) \\ &= 1 - \mathbb{P}(Y \ge y) \\ &= 1 - \mathbb{P}\left(\min\{w_{i1}, w_{i2}, \dots, w_{iN}\} \ge y\right) \\ &= 1 - \mathbb{P}\left(w_{i1} \ge y\right) \cdot \mathbb{P}\left(w_{i2} \ge y\right) \cdot \dots \cdot \mathbb{P}\left(w_{iN} \ge y\right) \\ &= 1 - e^{-\lambda_1 y} \cdot e^{-\lambda_2 y} \cdot \dots \cdot e^{-\lambda_N y} \\ &= 1 - e^{-\sum_{j=1}^{N} \lambda_j y} \qquad y > 0 \end{aligned} \tag{21}$$

The latter cumulative distribution function is that of an exponential variable with parameter $\sum_{j=1}^{N} \lambda_j$. □

Lemma 2. *For S_k defined as above we have* [15]*:*

$$0.67 \cdot \sqrt{k} \lessapprox \frac{2\pi}{e^2} \cdot \sqrt{\frac{2}{\pi}} \cdot \sqrt{k} \le \mathbb{E}\left[|S_k|\right] \le \frac{e}{\sqrt{\pi}} \cdot \sqrt{\frac{2}{\pi}} \cdot \sqrt{k} \lessapprox 1.23 \cdot \sqrt{k} \tag{22}$$

Proof. The starting point of our proof is an intermediate result in the proof of the limit of the expected absolute value of the 1-d random walk, which is detailed in Hizak and Logozar (2011, Equations 29a and 29b) and is based on combinatorial arguments via the binomial distribution:

$$\mathbb{E}[|S_k|] = \begin{cases} \dfrac{1}{2^{k-2}} \dfrac{k}{2} \dbinom{k-1}{k/2} = \dfrac{k}{2^k} \dfrac{k!}{[(k/2)!]^2} & \text{for } k \text{ even,} \\ \dfrac{1}{2^{k-1}} \dfrac{k+1}{2} \dbinom{k}{(k+1)/2} = \dfrac{k+1}{2^{k+1}} \dfrac{(k+1)!}{[((k+1)/2)!]^2} & \text{for } k \text{ odd.} \end{cases} \tag{23}$$

Since $\mathbb{E}[|S_{2k}|] = \mathbb{E}[|S_{2k-1}|]$ it suffices to analyze the case where k is even. To that end, we will use Stirling's formula to bound $k!$ from above and below as

$$\sqrt{2\pi} \cdot k^{k+1/2} \cdot e^{-k} \le k! \le e \cdot k^{k+1/2} \cdot e^{-k} \tag{24}$$

For k even, we may bound $|S_k|$ from above as:

$$\mathbb{E}[|S_k|] = \frac{k}{2^k} \frac{k!}{[(k/2)!]^2} \le \frac{k}{2^k} \frac{e \cdot k^{k+\frac{1}{2}} \cdot e^{-k}}{2\pi \cdot (k/2)^{k+1} \cdot e^{-k}} = \frac{e}{\sqrt{2\pi}} \cdot \sqrt{\frac{2}{\pi}} \cdot \sqrt{k} \tag{25}$$

[15] Note that $\lim_{k\to\infty} \mathbb{E}[|S_k|] = \sqrt{\frac{2}{\pi}} \cdot \sqrt{k}$ (Peters, 1856).

Next, we lower bound $|S_k|$ for k even:

$$\mathbb{E}[|S_k|] = \frac{k}{2^k}\frac{k!}{[(k/2)!]^2} \geq \frac{k}{2^k}\frac{\sqrt{2\pi}\cdot k^{k+1/2}\cdot e^{-k}}{e^2\cdot(k/2)^{k+1}\cdot e^{-k}} = \frac{2\pi}{e^2}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{k} \quad (26)$$

This concludes the proof for k even. For k odd we have with the observation that $|S_k| = |S_{k+1}|$:

$$\frac{2\pi}{e^2}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{k} \leq \frac{2\pi}{e^2}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{2\lceil k/2\rceil} \leq \mathbb{E}\left[|S_{k+1}|\right] = \mathbb{E}\left[|S_k|\right] \quad (27a)$$

and

$$\frac{e}{\sqrt{\pi}}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{k} \geq \frac{e}{\sqrt{\pi}}\cdot\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2}}\cdot\sqrt{2\lceil k/2\rceil} \geq \mathbb{E}\left[|S_{k+1}|\right] = \mathbb{E}\left[|S_k|\right] \quad \square \quad (27b)$$

Lemma 3. *The following relations between the stopping times hold:*

(i) $t(k,C) = t(1,C) + t(k-1,1)$
(ii) $t(1,C) = t(C,1)$

Proof. (i) $t(k,C)$ is the time at which for the kth time at least C clients and C providers are in the market, assuming that every time the latter is the case one client and one provider are removed. Now consider $t(1,C)$, this is the first time at least C clients and C providers are in the market and thus the first matching event. After that point there are always at least $C-1$ clients and providers (as a matching event occurs when there are at least C on each side and removes only one from each market side). Considering this as the stock, the remaining process is equivalent to the process where a matching event occurs every time at least one client and one provider are in the market. As another $k-1$ agents need to be matched, we have $t(k,C) = t(1,C) + t(k-1,1)$.
(ii) $t(1,C)$ is the first time at least C clients and C providers are in the market and no one has exited thus far. On the other hand, $t(C,1)$ is the Cth time at least one client and one provider were in the market and each time a matching event occurred one client and one provider were removed. The latter is equivalent to the former. □

Lemma 4. *The following bounds for the expected stopping times hold:*

(i) $\mathbb{E}[t(k,1)] < 5k$
(ii) $\mathbb{E}[t(k,C)] < 5C + 5k$

Proof. (i) By Lemma 2 we have $\mathbb{E}[S_t] < 1.23\sqrt{t}$. Thus, at time t, after the arrival of in expectation t agents, the expected difference of the number of agents that arrived to each market side is upper bounded by $1.23\sqrt{t}$., The expected number of agents that have arrived to the short side of the market by time t is then lower bounded by $\frac{t-1.23\sqrt{t}}{2}$. (Recall that for $t(k,1)$ every time when there is at least 1 client and 1 provider in the market, one pair is removed.)

Estimating $t(k,1)$ entails estimating the expected time it takes until k agents have arrived to the short side of the market. Thus we set $k = \frac{t-1.23\sqrt{t}}{2}$ and solve for t. Rearranging and using the variable transformation $t = u^2$ we have the quadratic equation:

$$u^2 - u - 2k \stackrel{!}{=} 0 \quad (28)$$

The solutions are:

$$u_{+,-} = \frac{1.23 \pm \sqrt{1.23^2 + 8k}}{2} \quad (29)$$

Given the variable transformation the positive solution is selected. Squaring u_+ we finally have:

$$\mathbb{E}[t(k,1)] = \Big(\frac{1.23 + \sqrt{1.23^2 + 8k}}{2}\Big)^2 < \frac{3}{4} + 2k + 2\sqrt{k} < 5k \quad (30)$$

(ii) Using (i) and Lemma 3 the result follows. □

Appendix B. Proofs of main results

B.1. Lemma 5

Lemma 5. *For exponentially distributed match costs, the matching cost of the patient clearing schedule* $\mathtt{CS}_{\mathtt{patient}}$ *is bounded for all* $A \geq 1$ *as follows:*

$$\frac{\log 2}{\bar{\lambda}} \leq \mathrm{cost}_{\mathtt{patient}}(A) \leq \frac{\pi^2}{6\underline{\lambda}}$$

Proof of Proposition 1. For the first assertion, note that the exponential distribution is closed under scaling by a positive factor, i.e., if $X \sim \exp(\kappa)$ then $\mu X \sim \exp(\kappa/\mu)$. In our case, this implies that

$$w_{ij} \sim \exp(\lambda_{ij}) \iff w_{ij} \sim \frac{1}{\lambda_{ij}}\exp(1) \quad (31)$$

We have that for all $i \in \mathcal{C}$, $j \in \mathcal{P}$, the distribution of w_{ij} is first-order stochastically dominated by $\underline{\lambda}^{-1}\exp(1)$. Thus the expected weight is upper bounded by the simplified problem where all match costs are distributed according to $\underline{\lambda}^{-1}\exp(1)$. With this in mind, we will simplify notation in the rest of the proof by setting $\underline{\lambda} = 1$.

By the summation formula of Buck et al. (2002) and Linusson and Waestlund (2004), we have for the expected weight of the minimum A-matching (note that $A = N$, recalling that $N = \min\{N_{\mathcal{C}}, N_{\mathcal{P}}\}$):

$$\mathbb{E}_{\min}\left[\sum_{k=1}^{N} w_{i_k,j_k}\right] = \sum_{\substack{i,j\geq 0\\ i+j<N}} \frac{1}{(N_{\mathcal{C}}-i)\cdot(N_{\mathcal{P}}-j)}. \quad (32)$$

Thus, we readily have

$$\mathrm{cost}_{\mathtt{patient}}(N) = \mathbb{E}_{\min}\left[\sum_{k=1}^{N} w_{i_k,j_k}\right] \leq \sum_{\substack{i,j\geq 0\\ i+j<N}} \frac{1}{(N-i)\cdot(N-j)}. \quad (33)$$

To proceed, by Waestlund (2009, Lemma 3.1) we have

$$\sum_{\substack{i,j\geq 0\\ i+j<N}} \frac{1}{(N-i)\cdot(N-j)} = \sum_{k=1}^{N}\frac{1}{k^2} \leq \zeta(2), \quad (34)$$

where $\zeta(2) = \sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$ is the Basel constant. Returning to our original problem, we conclude that $\mathrm{cost}_{\mathtt{patient}}(A) \leq \pi^2/(6\underline{\lambda})$, as claimed.

To compute the lower bound, we can consider the expected match cost for $A = 1$, because matching more players would only increase the expected matching cost. Let $Y \geq 2$ be the number of agents required such that at least one client and one provider have entered the market. Then the event $Y = k+1$ is the same event as the union of the disjoint events 'the first k agents are clients and the $(k+1)$th agent is a provider' and 'the first k agents are providers and the $(k+1)$th agent is a client'. Each of the latter events has probability $1/2^{k+1}$, so $\mathbb{P}(Y = k+1) = 2^{-k}$. Moreover, as we prove in Lemma 1, for $Y = k+1$, the expected minimum match cost is bounded below by $\frac{1}{\bar{\lambda}\cdot k}$. Thus for $A = 1$, we get

$$\mathrm{cost}_{\mathtt{patient}}(N) = \sum_{k=1}^{\infty}\frac{1}{\bar{\lambda}k}\mathbb{P}(Y = k+1) = \frac{1}{\bar{\lambda}}\sum_{k=1}^{\infty}\frac{1}{2^k k} = \frac{\log 2}{\bar{\lambda}}, \quad (35)$$

where the last equality follows from the series expansion $\log(1-x) = -x - x^2/2 - x^3/3 - \cdots$ applied to $x = 1/2$. □

B.2. Theorem 4

Theorem 4. *The expected matching cost ratios for the schedules under study are as follows:*

$(\mathtt{CS}_{\mathtt{FCFS}})$	*FCFS matching:*	$\alpha_{\mathtt{FCFS}} = \frac{6\bar{\lambda}\lambda}{\pi^2}A$	(36a)
$(\mathtt{CS}_{\mathtt{greedy}})$	*Greedy matching:*	$\alpha_{\mathtt{greedy}} \geq \frac{6\underline{\lambda}}{5\pi^2}A^{1/2}$	(36b)
$(\mathtt{CS}_{0\leq\gamma<1/2})$	*Subcritical rate matching:*	$\underline{C}_\gamma A^{1/2-\gamma} \leq \alpha_{0\leq\gamma<1/2} \leq \bar{C}_\gamma A^{1-2\gamma}$	(36c)

$(\mathtt{CS}_{1/2})$ *Critical rate matching:* $\frac{2\underline{\lambda}}{\pi^2}\log A \leq \alpha_{1/2} \leq \frac{\overline{\lambda}}{\underline{\lambda}}\frac{1+\log A}{\log 2}$ (36d)

$(\mathtt{CS}_{1/2<\gamma\leq 1})$ *Supercritical rate matching:* $\alpha_{1/2<\gamma\leq 1} = \frac{\overline{\lambda}}{\underline{\lambda}}\frac{\zeta(2\gamma)}{\log 2}$ (36e)

$(\mathtt{CS}_{\mathtt{patient}})$ *Patient matching:* $\alpha_{\mathtt{patient}} = 1$ (36f)

Remark. In the above, $\underline{C}_\gamma$ and $\overline{C}_\gamma$ are positive constants, and $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ denotes the Riemann zeta function (so $\zeta(s) < \infty$ for all $s > 1$).

It is worth noting that the bounds for α become asymptotically 'less tight' for small $\gamma < \frac{1}{2}$. As far as this gap is concerned, we conjecture that the upper bound is the tight one: the lower bound is obtained via a crude approximation using Jensen's inequality (Jensen, 1906), and this could be potentially tightened (although we have not been able to do so). By contrast, the approximation for the upper bound seems less drastic.

For the sake of limiting notations the proposition and proof are stated for the clearing schedules with $f(k) = \lceil k^\gamma \rceil$ rather than for $\Theta(f)$. Adding constant upper and lower bounds is straightforward and thus omitted. Recall that $t(k, f(k))$ is the stopping time for the event that for the kth time at least $f(k)$ clients and $f(k)$ providers are in the market, assuming that every time this is the case one client and one provider are removed.

Proof of Theorem 4. Throughout the proof we shall simplify notation by omitting the fact that some of the matching schedules are defined via the ceiling of functions mapping to $\mathbb{R}^+$ (e.g., $\lceil k^\gamma \rceil$). The results are not changed by the omission since match costs are never underestimated and overestimated by very little. Further, while they are stated together in the proposition, we study the clearing schedules $\mathtt{CS}_0$ and $\mathtt{CS}_{0<\gamma<1/2}$ separately since they require different arguments.

First come, first served $(\mathtt{CS}_{\mathtt{FCFS}})$ In FCFS the cost of each match is the expectation of a single match cost, that is λ. After A matches have occurred, the expected incurred cost is λA. Thus, given the patient clearing schedule has cost bounded above by $\frac{\pi^2}{6\underline{\lambda}}$, the expected matching cost ratio is equal to $\frac{\lambda \cdot A}{\pi^2/(6\underline{\lambda})}$.

Before stating the proofs for the other results recall from the proof of Proposition 1, that the exponential distribution is closed under scaling. We shall thus simplify notation and assume that for all i, j $w_{ij} \sim \exp(1)$. Note that, for lower bounds the scaling factor $\frac{1}{\overline{\lambda}}$ needs to be applied and for upper bounds the scaling $\frac{1}{\underline{\lambda}}$ needs to be applied. But note that those scaling factors are constant with respect to τ (and thus A) and therefore do not influence the orders of the limiting results.

Greedy matching $(\mathtt{CS}_{\mathtt{greedy}})$ The kth match happens when the minimum of the number of clients and providers who already arrived to the market is k, that is, at time $t(k, 1)$. The expected weight of the kth match depends on the number of players currently present on the long side of the market (since on the short side there is only one agent). This random variable is given by $|S_{t(k,1)}| + 1$. By Lemma 1 the expected weight thus is $\mathbb{E}[\frac{1}{|S_{t(k,1)}|+1}]$. The first A matches thus have an expected cost of

$$\mathbb{E}[\sum_{k=1}^{A} \frac{1}{|S_{t(k,1)}| + 1}]. \tag{37}$$

Given that we study fixed A (the number of matches that happened) we have[16]:

$$\mathbb{E}\left[\sum_{k=1}^{A} \frac{1}{|S_{t(k,1)}| + 1}\right] = \sum_{k=1}^{A} \mathbb{E}[\frac{1}{|S_{t(k,1)}| + 1}] \tag{38}$$

Next, by Jensen's inequality (Jensen, 1906; $\frac{1}{x}$ is convex) we have:

$$\sum_{k=1}^{A} \mathbb{E}[\frac{1}{|S_{t(k,1)}| + 1}] > \sum_{k=1}^{A} \frac{1}{\mathbb{E}[|S_{t(k,1)}| + 1]} = \sum_{k=1}^{A} \frac{1}{\mathbb{E}[|S_{t(k,1)}|] + 1} \tag{39}$$

By Lemma 2 we have $\mathbb{E}[|S_t|] < 1.23\sqrt{t}$ and by Lemma 4 $\mathbb{E}[t(k,1)] < 5k$, thus:

$$\sum_{k=1}^{A} \frac{1}{\mathbb{E}[|S_{t(k,1)}|] + 1} > \sum_{k=1}^{A} \frac{1}{\mathbb{E}[|S_{5k}|] + 1} > \frac{1}{1.23}\sum_{k=1}^{A} \frac{1}{\sqrt{5k}+1} \tag{40}$$

$$> \frac{1}{1.23}A \cdot \frac{1}{\sqrt{5A}+1} > A \cdot \frac{1}{5\sqrt{A}} = \frac{\sqrt{A}}{5} \tag{41}$$

Thus, given the optimal schedule has $\mathrm{cost}_{\mathtt{patient}}(A) \leq \frac{\pi^2}{6\underline{\lambda}}$, the expected matching cost ratio is lower bounded by $\frac{\sqrt{A}}{5\pi^2/(6\underline{\lambda})}$.

Subcritical matching $(\mathtt{CS}_0)$ We shall fix the clearing schedule such that it matches a couple every time some fixed $C \in \mathbb{N}$ players are on the short side of the market $(N - A = C)$ and note that it belongs to the family of clearing schedules $\mathtt{CS}_0$.

Next, by Lemma 4, $\mathbb{E}[t(k, C)] < 5C + 5k$. With the latter and, as above by Jensen's inequality (Jensen, 1906; $\frac{1}{x}$ is convex) and Lemma 2 we have for the expected matching cost:

$$\begin{aligned}
\mathbb{E}\sum_{k=1}^{A} \frac{1}{(C + |S_{t(k,C)}|)C} &\geq \sum_{k=1}^{A} \frac{1}{(C + \mathbb{E}[|S_{t(k,C)}|])C} \geq \sum_{k=1}^{A} \frac{1}{(C + \mathbb{E}[|S_{5C+5k}|])C} \\
&\geq \frac{1}{1.23}\sum_{k=1}^{A} \frac{1}{(C + \sqrt{5C+5k})C} \\
&\geq \frac{A}{1.23} \cdot \frac{1}{(C + \sqrt{5C+5A})C} \\
&= \frac{1}{1.23 \cdot C} \cdot \frac{A - C}{\sqrt{5C+5A} + C} = \Omega(\sqrt{A})
\end{aligned} \tag{42}$$

Thus, given the optimal clearing schedule has finite cost the expected matching cost ratio is $\alpha(A) = \Omega(\sqrt{A})$.

The second part of the assertion follows by observing:

$$\mathbb{E}[\sum_{k=1}^{A} \frac{1}{(C + |S_{t(k,C)}|)C}] < \frac{1}{C}\mathbb{E}[\sum_{k=1}^{A} \frac{1}{1 + |S_{t(k,1)}|}] \tag{43}$$

Subcritical matching $(\mathtt{CS}_{0<\gamma<1/2})$ For the upper bound, by Lemma 1, we have:

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{A} \frac{1}{(k^\gamma + |S_{t(k,\sqrt{k})}|)k^\gamma}\right] &\leq \sum_{k=1}^{A} \frac{1}{k^{2\gamma}} = 1 + \sum_{k=2}^{A} \frac{1}{k^{2\gamma}} \leq 1 + \int_{x=1}^{A} \frac{1}{x^{2\gamma}}\,dx \\
&= 1 + \left[\frac{1}{1-2\gamma}x^{1-2\gamma}\right]_{x=1}^{A} \leq 1 + \frac{1}{1-2\gamma}A^{1-2\gamma}
\end{aligned} \tag{44}$$

Thus, given the optimal clearing schedule has finite cost, the expected matching cost ratio is $\alpha(A) = \mathcal{O}(A^{1-2\gamma})$ for $0 < \gamma < \frac{1}{2}$.

For the lower bound, note that $t(k, k^\gamma) < 10k$ for $\gamma < 1$ by Lemma 4. Further note that $\mathbb{E}[|S_t|]$ is strictly increasing in t. Thus, with Jensen's inequality (Jensen, 1906; $\frac{1}{x}$ is convex):

$$\begin{aligned}
\sum_{k=1}^{A} \mathbb{E}[\frac{1}{(k^\gamma + |S_{t(k,k^\gamma)}|)k^\gamma}] &> \sum_{k=1}^{A} \frac{1}{(k^\gamma + \mathbb{E}[|S_{10k}|])k^\gamma} \\
&> \frac{1}{1.23}\sum_{k=1}^{A} \frac{1}{(k^\gamma + \sqrt{10k})k^\gamma} \\
&> \frac{1}{1.23(\sqrt{10}+1)}\sum_{k=1}^{A} \frac{1}{k^{\frac{1}{2}+\gamma}} \\
&> \frac{1}{6}\int_{x=1}^{A} \frac{1}{x^{\frac{1}{2}+\gamma}}dx \\
&> \frac{1}{6}[\frac{1}{\frac{1}{2}-\gamma}x^{\frac{1}{2}-\gamma}]_{x=1}^{A} = \Omega(A^{\frac{1}{2}-\gamma})
\end{aligned} \tag{45}$$

[16] Note that t (the total number of client and providers who have arrived to the market) depends on A (and vice versa). Therefore, Wald (1944)'s equation does not apply and thus the route of inquiry to study the matching cost at some continuous time τ does not work since we could not interchange summation and expectation.

Critical matching ($CS_{1/2}$) For the upper bound, by Lemma 1, we have:

$$\mathbb{E}[\sum_{k=1}^{A} \frac{1}{(\sqrt{k}+|S_{t(k,\sqrt{k})}|)\sqrt{k}}] < \sum_{k=1}^{A} \frac{1}{k} \leq \log A + 1 \tag{46}$$

where the last inequality follows from the bounds for the harmonic series. Thus, given the optimal clearing schedule has finite matching cost (lower bounded by $\frac{\log(2)}{\bar{\lambda}}$), the expected matching cost ratio is smaller than $\frac{\frac{1}{\underline{\lambda}}(\log A+1)}{\log(2)/\bar{\lambda}}$.

For the lower bound note that $\mathbb{E}[t(k,\sqrt{k})] < 10k$ by Lemma 4. Further note that S_t is strictly increasing in t. Thus, with Jensen's inequality (Jensen, 1906; $\frac{1}{x}$ is convex) and Lemma 2:

$$\begin{aligned}\sum_{k=1}^{A} \mathbb{E}[\frac{1}{(\sqrt{k}+|S_{t(k,\sqrt{k})}|)\sqrt{k}}] &> \sum_{k=1}^{A} \frac{1}{(\sqrt{k}+\mathbb{E}[|S_{10k}|])\sqrt{k}} \\ &> \frac{1}{1.23}\sum_{k=1}^{A} \frac{1}{(\sqrt{k}+\sqrt{10k})\sqrt{k}} \\ &\geq \frac{1}{6}\sum_{k=1}^{A}\frac{1}{k} > \frac{1}{6}\log A\end{aligned} \tag{47}$$

Thus, given the optimal clearing schedule has $\mathrm{cost}_{\mathtt{patient}}(A) \leq \frac{\pi^2}{6\underline{\lambda}}$, the expected matching cost ratio is bounded below by $\frac{\frac{1}{6}\log A}{\pi^2/(6\underline{\lambda})} = \frac{\underline{\lambda}}{\pi^2}\cdot\log A$.

Supercritical matching ($CS_{1/2<\gamma\leq 1}$) As above, by Jensen's inequality (Jensen, 1906; since $1/x$ is convex) and Lemma 1 we have:

$$\mathbb{E}[\sum_{k=1}^{A} \frac{1}{(k^\gamma+|S_{t(k,k^\gamma)}|)k^\gamma}] < \sum_{k=1}^{A}\mathbb{E}[\frac{1}{k^\gamma k^\gamma}] = \sum_{k=1}^{A}\frac{1}{k^{2\gamma}} \to \zeta(2\gamma) \tag{48}$$

where ζ is the Riemann zeta function and is known to converge for $\gamma > \frac{1}{2}$. Given that we are considering a sum with positive summands convergence is from below. Thus, given the optimal clearing schedule has finitematching cost ($\mathrm{cost}_{\mathtt{patient}}(A) \geq \frac{\log(2)}{\bar{\lambda}}$), the expected matching cost ratio is bounded from above by $(\bar{\lambda}/\underline{\lambda})\cdot\zeta(2\gamma)/\log 2$ for $\frac{1}{2} < \gamma \leq 1$. □

B.3. Theorem 5

Theorem 5. *The expected waiting time ratios for the schedules under study are as follows:*

($CS_{\mathtt{FCFS}}$)	*FCFS matching:*	$\beta_{\mathtt{FCFS}} = 1$	(49a)
($CS_{\mathtt{greedy}}$)	*Greedy matching:*	$\beta_{\mathtt{greedy}} = 1$	(49b)
($CS_{0\leq\gamma<1/2}$)	*Subcritical rate matching:*	$\beta_{0\leq\gamma<1/2} = \Theta(1)$	(49c)
($CS_{1/2}$)	*Critical rate matching:*	$\beta_{1/2} = \Theta(1)$	(49d)
($CS_{1/2<\gamma\leq 1}$)	*Supercritical rate matching:*	$\beta_{1/2<\gamma\leq 1} = \Theta(\tau^{\gamma-1/2})$	(49e)
($CS_{\mathtt{patient}}$)	*Patient matching:*	$\beta_{\mathtt{patient}} = \Theta(\tau^{1/2})$	(49f)

The results in Theorem 5 are driven by the assumption that the arrival of either a client or a provider at every stage of the process is equally likely. This entails that the expected absolute difference of clients and providers $|N_{\mathcal{C}}(\tau) - N_{\mathcal{P}}(\tau)|$ can by approximated by a Wiener process as detailed in the proof of Proposition 1. For the latter we know that the expectation is $\sqrt{\tau}$, so $|N_{\mathcal{C}}(\tau) - N_{\mathcal{P}}(\tau)| \approx \sqrt{\tau}$ in expectation.

Proof of Theorem 5. First note that to compare different clearing schedules we are interested in the additional waiting time incurred until some number A of couples have been matched. Thus, we consider the waiting time until T for the greedy schedule (the benchmark) and for other schedules the waiting time until $\hat{T}$ where $\hat{T}$ is the expected time until under the given schedule the same number of couples have been matched as in the greedy schedule until time T.

As in the Proof of Theorem 4 we shall simplify notation by omitting the fact that some of the matching schedules are defined via the ceiling function of functions mapping to $\mathbb{R}^+$ (e.g., $\lceil k^\gamma \rceil$). We invite the reader to convince her-or himself that the results are not altered through this simplification.

Let $\tau(k)$ be the moment the kth couple is matched (given a particular clearing schedule). We proceed in a case-by-case basis below:

First come, first served ($CS_{\mathtt{FCFS}}$) It suffices to note that this clearing schedules matches players at exactly the same moments as the greedy clearing schedules. The result then follows.

Subcritical and critical matching ($CS_{0\leq\gamma\leq 1/2}$) We shall study the worst case such clearing schedule with respect to waiting time. We consider two different parts. In the first part we wait until at least T^γ clients and T^γ providers are in the market. The second part then proceeds in the same way as the greedy clearing schedule, keeping in mind that at all future times $\min\{N_{\mathcal{C}}, N_{\mathcal{P}}\}$ is exactly T^γ. The expected waiting time of the first schedule can be bounded above by the upper bound for the expected time until T^γ clients and T^γ providers are in the market, that is, $\mathbb{E}[\tau(5T^\gamma)] = 5T^\gamma$ (see Lemma 4) noting that we used the fact that the arrival of agents is governed by a Poisson clock of rate 1. Now, a crude upper bound for the waiting time of the first part of the process is found be assuming that all agents are in the market from the beginning ($\tau = 0$), yielding the upper bound $5T^\gamma \cdot 5T^\gamma$.

Note that, the first part of the process takes $\hat{T} - T$ time. For the remaining second part of the process the waiting time is the time of the greedy schedule ($\Theta(T^{3/2})$) plus the cost of the – in expectation – no more than $5T^\gamma$ agents on each side of the market to 'remain' for the subsequent periods. Thus the total waiting time is bounded above by:

$$5T^\gamma \cdot 5T^\gamma + \frac{2}{3}T^{3/2} + 5T^\gamma \cdot T = \Theta(T^{3/2}) \tag{50}$$

Thus $\beta(\hat{T}) = (3/2)\,\Theta(T^{3/2})/\Theta(T^{3/2}) = \Theta(1)$.

Supercritical matching ($CS_{1/2<\gamma\leq 1}$) We first construct a lower bound. Consider the alternative arrival process, where clients and providers alternatingly arrive to the market. Note that for any given clearing schedule this process incurs lower waiting time. For the clearing schedule we consider the waiting time of this alternative arrival process is precisely governed by the fact that the kth match takes place when at least k^γ players are on the short side of the market. Further note that $\hat{T} \geq T$. Thus, the waiting time is lower bounded by using the approximation by the Wiener process (by arguments as in Proposition 1 and by observing that arrival is governed by a Poisson clock of rate 1):

$$\int_0^T 2\tau^\gamma d\tau = \frac{2}{1+\gamma}\tau^{1+\gamma}|_0^T = \Omega(T^{1+\gamma}) \tag{51}$$

For the upper bound, we construct a clearing schedule that constitutes an upper bound of the schedule under consideration. For fixed k, let $T = \tau(k)$ consider the following clearing schedule: First wait until there are at least k^γ clients and providers in the market, then proceed with the greedy schedule such that at any future point $\min\{clients, providers\}$ in the market is equal to k^γ. Note that this new schedule has the same total run time as the original schedule, that is, $\hat{T}$. Further it is evident that the waiting time occurred by the new schedule is greater than the waiting time of the original schedule. By arguments as for (CS_0) and by the fact that arrival is governed by a Poisson clock of rate 1 we can upper bound the waiting time by:

$$5T^\gamma \cdot 5T^\gamma + \frac{2}{3}T^{3/2} + 5T^\gamma \cdot T = \Theta(T^{1+\gamma}) \tag{52}$$

since we assumed $\frac{1}{2} < \gamma \leq 1$.

The two bounds together show that the waiting time of the originally considered clearing schedule is $\Theta(T^{1+\gamma})$. Thus $\beta(\hat{T}) = \frac{\Theta(T^{1+\gamma})}{\Theta(T^{3/2})} = \Theta(T^{\gamma-\frac{1}{2}})$.

Patient matching ($CS_{\mathtt{patient}}$) First note that for the patient schedule $\hat{T} = T$. The expected waiting time for the patient schedule until time T

is given by

$$\mathbb{E}\left[\int_0^T N_{\mathcal{C}}(\tau) + N_{\mathcal{P}}(\tau)\, d\tau\right] = \int_0^T \mathbb{E}\left[N_{\mathcal{C}}(\tau) + N_{\mathcal{P}}(\tau)\right]\, d\tau \tag{53}$$

where the latter equality holds by Tonelli (1909)'s theorem (by noting that $N_{\mathcal{C}}(\tau) + N_{\mathcal{P}}(\tau)$ is non-negative). The expectation is with respect to the number of clients and providers and with respect to the arrival times of the agents (governed by a Poisson clock). Again by Tonelli (1909)'s theorem we can consider the case where the expectation with respect to the arrival times is taken first. Then by the fact that the arrival of agents is assumed to follow a Poisson clock of rate 1 we have:

$$\int_0^T \mathbb{E}[N_{\mathcal{C}}(\tau) + N_{\mathcal{P}}(\tau)] d\tau = \int_0^T \lfloor \tau \rfloor d\tau = \Theta(T^2) \tag{54}$$

Thus $\beta(\hat{T}) = \frac{\Theta(T^2)}{\Theta(T^{3/2})} = \Theta(\sqrt{T})$. □

B.4. Proof of Theorem 3

Proof of Theorem 3. Consider a clearing schedule of the form $\mathtt{CS}_f$ that matches the kth couple when $\lceil f(k)\rceil$ players on the short side of the market. To balance the expected matching cost and waiting time ratios, any such clearing schedule would have to satisfy $f(k) = \omega(\sqrt{k})$; otherwise, the expected matching cost ratio would dominate asymptotically the expected waiting time ratio (see Table 2).

Recall that $t(k, f(k))$ is the stopping time for the event that for the kth time at least $f(k)$ clients and $f(k)$ providers are in the market, assuming that every time this is the case one client and one provider are removed. Further, recall that $S_\tau = N_{\mathcal{C}}(\tau) - N_{\mathcal{P}}(\tau)$ is the difference of clients and providers who have arrived to the market until τ.

We begin with the expected matching cost ratio. For the upper bound we have:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=1}^{A} \frac{1}{(\lceil f(k)\rceil + |S_{t(k,f(k))}|)\lceil f(k)\rceil}\right] \\
&= \sum_{k=1}^{A} \mathbb{E}\left[\frac{1}{(\lceil f(k)\rceil + |S_{t(k,f(k))}|)\lceil f(k)\rceil}\right] \\
&\le \sum_{k=1}^{A} \lceil f(k)\rceil^{-2}
\end{aligned} \tag{55}$$

For the lower bound, we show in Appendix A, Lemma 4 that $\mathbb{E}[t(k, f(k))] < 10k$. Furthermore, note that $\mathbb{E}\left[|S_t|\right]$ is strictly increasing in t. Thus, by Jensen (1906)'s inequality, and Lemma 2 (which is bounding $|S_t|$ via a combinatorial argument and using Stirling's formula), we get:

$$\begin{aligned}
\sum_{k=1}^{A} \mathbb{E}\left[\frac{1}{(\lceil f(k)\rceil + |S_{t(k,f(k))}|)\lceil f(k)\rceil}\right] &\ge \sum_{k=1}^{A} \frac{1}{(\lceil f(k)\rceil + \mathbb{E}[|S_{10k}|])\lceil f(k)\rceil} \\
&\ge \frac{\pi}{\sqrt{2e}} \sum_{k=1}^{A} \frac{1}{(\lceil f(k)\rceil + \sqrt{10k})\lceil f(k)\rceil} \\
&= \Theta\left(\sum_{k=1}^{A} \frac{1}{f(k)^2}\right)
\end{aligned} \tag{56}$$

where the last line follows from the assumption $f(k) = \omega(k^{1/2})$. Thus the two bounds together with the fact that the patient schedule has finite matching cost yield the result that, for $f(k) = \omega(k^{1/2})$ the expected matching cost ratio is $\alpha(A) = \Theta\left(\sum_{k=1}^{A} 1/\left[f(k)^2\right]\right)$.

We proceed, by considering the incurred waiting time. Recall that $N(\tau) - A(\tau)$ is the number of agents on the shorter side of the market at time τ, so $N - A = \lceil f(k-1)\rceil - 1$ after the $(k-1)$st match. The number of clients and the number of providers that need to arrive to the market before the kth match is thus upper bounded by

$$\lceil f(k)\rceil - \left(\lceil f(k-1)\rceil - 1\right) = 1 + \lceil f(k)\rceil - \lceil f(k-1)\rceil \le 2 \tag{57}$$

where the last inequality follows from the fact that $f(k) = o(k)$ is a necessary condition for a feasible clearing schedule (that is, a clearing schedule where the proportion of unmatched versus matched players is decreasing). This is the case as for a clearing schedule with $f(k) = \Omega(k)$ the kth couple is matched when more than k providers and k clients are in the market. Thus the number of remaining agents in the market is $\Omega(k)$ and thus the ratio of matched versus unmatched agents (among agents who entered the market) does not tend to 1. The expected waiting time accrued between the $(k-1)$st and the kth match is therefore upper bounded by:

$$\Delta_k := \underbrace{\mathbb{E}[\text{time s.t.} \ge 2 \text{ clients } \& \ge 2 \text{ providers enter market}]}_{=:\Delta_k^1} \cdot \underbrace{\left(2\lceil f(k)\rceil + \lceil g(k)\rceil\right)}_{=:\Delta_k^2} \tag{58}$$

where $\lceil g(k)\rceil$ is a function that we will use to upper bound $|S_k|$, viz., the random variable constituting the absolute difference of clients and providers in the market at time $\tau(k)$. For posterity, note also that Δ_k^1 is the expectation of the time between the $(k-1)$th and the kth match and Δ_k^2 provides an upper bound for the number of agents waiting in the time interval between the $(k-1)$th and the kth match.

Given the arrival of agents is governed by a Poisson clock of rate one, we have $\Delta_k^1 = 5$, i.e., on average, five agents need to enter the market to have at least two clients and at least two providers. To see this, let Y be the number of flips of a coin required to observe at least 2 heads (clients) and 2 tails (providers). The event '$Y > k$' is then equivalent to the union of the events '$\binom{k}{k-1}$ heads' and '$\binom{k}{k-1}$ tails'. The two latter events are disjoint and each has probability $\frac{k}{2^k}$. Thus $\mathbb{P}[Y > k] = \frac{k}{2^{k-1}}$ and we have

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} \mathbb{P}(Y > k) = 1 + 2\sum_{k=1}^{\infty} \frac{k}{2^k} = 1 + 2\sum_{k=1}^{\infty}\sum_{j=1}^{k} \frac{1}{2^k} \tag{59}$$

$$= 1 + 2\sum_{j=1}^{\infty}\sum_{k=j}^{\infty} \frac{1}{2^k} = 1 + 2\sum_{j=1}^{\infty} \frac{1}{2^{j-1}} = 5 \tag{60}$$

We thus have for Eq. (58)

$$\Delta_k = 5 \cdot \left(2\lceil f(k)\rceil + \lceil g(k)\rceil\right) \tag{61}$$

Next, to choose the function $g(k)$, note that the law of the iterated logarithm (Khintchine, 1924) gives

$$\limsup_{k\to\infty} \frac{|S_k|}{\sqrt{2k \log\log k}} = 1. \tag{62}$$

Hence, by choosing $g(k) = \sqrt{2k \log\log k}$, the random variable $|S_k|$ is asymptotically bounded from above by $g(k)$ with probability one.

We consider two cases below, which are exhaustive by Hardy (1910, Theorem, page 18):

Case 1: $f(k) = \Omega(g(k))$. For the first case we have:

$$5 \cdot (2f(k) + g(k)) = \Theta(f(k)) \tag{63}$$

The expected waiting time ratio until A pairs have been matched is bounded from above by $A^{-3/2}\sum_{k=1}^{A} \Theta(f(k))$, where we are using the fact that the expected waiting time for the greedy schedule is given by $\Theta(A^{3/2})$ (see Proposition 1). A trivial lower bound for the expected waiting time ratio is then given by

$$\frac{1}{A^{3/2}}\sum_{k=1}^{A} 2f(k) = \frac{1}{A^{3/2}}\sum_{k=1}^{A} \Theta(f(k)) \tag{64}$$

Thus, the expected waiting time ratio is given by

$$\beta(A) = \Theta\left(\frac{1}{A^{3/2}}\sum_{k=1}^{A} f(k)\right) \tag{65}$$

Moving to the comparison of matching cost and waiting time ratios, we recall that u_{cost} and u_{wait} are decreasing and concave and are of the same order (by assumption). Thus $u = u_{\text{cost}} + u_{\text{wait}}$ is maximized if and only if $\alpha = \Theta(\beta)$. In turn, this holds if and only if

$$\sum_{k=1}^{A} \frac{1}{f(k)^2} = \Theta\left(\frac{1}{A^{3/2}} \sum_{k=1}^{A} f(k)\right) \tag{66}$$

Recalling that f is assumed non-decreasing for large k, the summand on the left-hand side is decreasing and

$$\int_0^A \frac{1}{f(x)^2}\, dx \geq \sum_{k=1}^{A} \frac{1}{f(k)^2} \geq \int_1^{A+1} \frac{dx}{f(x)^2} \tag{67}$$

Considering the meaning of $f(k)$ it is without loss of generality to define $f(x) = 1$ for $x \in [0, 1)$ since the summand $\frac{1}{f(k)^2}$ remains decreasing. Thus the absolute difference between the two bounds is bounded above by:

$$\begin{aligned} &\left|\int_0^A \frac{1}{f(x)^2}\, dx - \int_1^{A+1} \frac{1}{f(x)^2}\, dx\right| \\ &= \left|\int_0^1 \frac{1}{f(x)^2}\, dx - \int_A^{A+1} \frac{1}{f(x)^2}\, dx\right| \leq 1 \end{aligned} \tag{68}$$

It follows that

$$\sum_{k=1}^{A} \frac{1}{f(k)^2} = \Theta\left(\int_1^{A+1} \frac{1}{f(x)^2}\, dx\right) \tag{69}$$

Next consider the right-hand side of (66). The summand is increasing, so we get:

$$\frac{1}{A^{3/2}} \int_0^A f(x)\, dx \leq \frac{1}{A^{3/2}} \sum_{k=1}^{A} f(k) \leq \frac{1}{A^{3/2}} \int_1^{A+1} f(x)\, dx \tag{70}$$

Now note that $f(x) < x$ must hold. Thus the absolute difference between the two bounds is bounded above by:

$$\begin{aligned} &\frac{1}{A^{3/2}} \left|\int_0^A f(x)\, dx - \int_1^{A+1} f(x)\, dx\right| \\ &= \frac{1}{A^{3/2}} \left|\int_A^{A+1} f(x) dx - \int_0^1 f(x)\, dx\right| \\ &\leq \frac{A}{A^{3/2}} = \mathcal{O}(1). \end{aligned} \tag{71}$$

It follows that

$$\frac{1}{A^{3/2}} \sum_{k=1}^{A} f(k) = \Theta\left(\frac{1}{A^{3/2}} \int_1^{A+1} f(x)\, dx\right) \tag{72}$$

With above approximations it follows that Eq. (66) holds if and only if the following equation holds:

$$\int_1^A \frac{1}{f(x)^2}\, dx = \Theta\left(\frac{1}{A^{3/2}} \int_1^A f(x)\, dx\right) \tag{73}$$

We shall show that $f(x) = \Theta(\sqrt{x}(\log x)^{1/3})$ is the unique solution to Eq. (73) up to order. To simplify notation, let $f(x) = \sqrt{x}(\log x)^{1/3}$, so the left-hand side (LHS) of Eq. (73) becomes

$$\int_1^A \frac{1}{x(\log x)^{2/3}} = 3(\log A)^{1/3} + c \tag{74}$$

where c is uniformly bounded and independent of A. Next, focusing on the RHS of Eq. (73), we get

$$\begin{aligned} \frac{1}{A^{3/2}} \int_1^A \sqrt{x}(\log x)^{1/3}\, dx &= (\log A)^{1/3} - \frac{1}{A^{3/2}} \int_1^A x^{3/2} \frac{1}{3x(\log x)^{2/3}}\, dx \\ &= (\log A)^{1/3} - \frac{1}{A^{3/2}} \underbrace{\int_1^A \sqrt{x} \frac{1}{3(\log x)^{2/3}}\, dx}_{=o\left(\int_1^A \sqrt{x}\cdot(\log x)^{1/3}\, dx\right)} \\ &= \Theta((\log A)^{1/3}) \end{aligned} \tag{75}$$

Our uniqueness claim follows by noting that the LHS of Eq. (73) is decreasing in $f(x)$ (in orders of magnitude of the upper bound of the integral) while the right-hand side (RHS) is increasing in $f(x)$.

Case 2: $f(k) = o(g(k))$. For the second case, assume that $f(k) = o(g(k))$. This implies for the matching cost that[17]

$$\int_1^A \frac{1}{f(\tau)^2}\, d\tau = \omega\left(\int_1^A \frac{1}{g(\tau)^2}\, d\tau\right) \tag{76}$$

The integral on the RHS of Eq. (76) can then be bounded from below as follows

$$\int_1^A \frac{1}{g(\tau)^2}\, d\tau = \int_1^A \frac{1}{4\tau \log\log \tau}\, d\tau = \omega\left(\int_1^A \frac{1}{4\tau(\log \tau)^{2/3}}\, d\tau\right) \tag{77}$$

For the integral on the RHS of Eq. (77) we have

$$\int_1^A \frac{1}{4\tau(\log \tau)^{2/3}} d\tau = \Theta\big((\log A)^{1/3}\big), \tag{78}$$

Hence, combining these last approximations, we finally get

$$\int_1^A \frac{1}{f(\tau)^2}\, d\tau = \omega\big((\log A)^{1/3}\big). \tag{79}$$

Thus any solution satisfying $f(k) = o(g(k))$ (Case 2) has expected matching cost that is $\omega(1)$ relative to the optimal solution. This completes the proof that $f(k) = \Theta(\sqrt{x}(\log(x+1))^{1/3})$ is the unique optimal clearing schedules for the balanced social planner by noting that we add 1 in the argument of the logarithm to initialize the clearing schedule for $k = 1$. □

[17] Formally, for $\tau \geq e$, the integrand is not well-defined, but the Cauchy principal value of the integral remains finite, and this is the value we are using for $\tau \leq e$. This issue could be side-stepped by shifting the lower limit of the integral to a higher value, but we do not do so to simplify the presentation.

References

Aggarwal, G., Goel, G., Karande, C., Mehta, A., 2011. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1253–1264.

Akbarpour, M., Li, S., Oveis Gharan, S., 2020. Thickness and information in dynamic matching markets. J. Polit. Econ. 128 (3), 783–815.

Aldous, D.J., 2001. The $\zeta(2)$ limit in the random assignment problem. Random Struct. Algorithms 18, 381–418.

Anderson, R., Ashlagi, I., Gamarnik, D., Kanoria, Y., 2015. A dynamic model of barter exchange. In: Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1925–1933.

Ashlagi, I., Azar, Y., Charikar, M., Chiplunkar, A., Geri, O., Kaplan, H., Makhijani, R., Wang, Y., Wattenhofer, R., 2017. Min-cost bipartite perfect matching with delays. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. APPROX/RANDOM 2017, Vol. 81, pp. 1–20.

Ashlagi, I., Burq, M., Jaillet, P., Manshadi, V.H., 2019. On matching and thickness in heterogeneous dynamic markets. Oper. Res. 67 (4), 927–949.

Ashlagi, I., Nikzad, A., Strack, P., 2023. Matching in dynamic imbalanced markets. Rev. Econom. Stud. 90 (3), 1084–1124.

Azar, Y., Chiplunkar, A., Kaplan, H., 2017. Polylogarithmic bounds on the competitiveness of min-cost perfect matching with delays. In: Proceedings of the Twenty-Eight Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1051–1061.

Baccara, M., Lee, S., Yariv, L., 2020. Optimal dynamic matching. Theor. Econ. 50, 1221–1278.

Bäumler, J., Bullinger, M., Kober, S., Zhu, D., 2023. Superiority of instantaneous decisions in thin dynamic matching markets. In: Proceedings of the 24th ACM Conference on Economics and Computation. pp. 390–390.

Blanchet, J.H., Reiman, M.I., Shah, V., Wein, L.M., Wu, L., 2022. Asymptotically optimal control of a centralized dynamic matching market with general utilities. Oper. Res. 70 (6), 3355–3370.

Bloch, F., Houy, N., 2012. Optimal assignment of durable objects to successive agents. Econom. Theory 51 (1), 13–33.

Buck, M.W., Chan, C.S., Robbins, D.P., 2002. On the expected value of the minimum assignment. Random Struct. Algorithms 21, 33–58.

Dow, J.C.R., Dicks-Mireaux, L.A., 1958. The excess demand for labour. a study of conditions in great britain, 1946-56. Oxf. Econ. Pap. 10 (1), 1–33.

Edmonds, J., 1965. Paths, trees, and flowers. Canad. J. Math. 17, 49–467.

Egervary, E., 1931. Matrixok kombinatorius tulajdonsagairol. Mat. Fiz. Lapok 38, 16–27.

Emek, Y., Kutten, S., Wattenhofer, R., 2016. Online matching: Haste makes waste!. In: Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing. pp. 333–344.

Emek, Y., Shapiro, Y., Wang, Y., 2019. Minimum cost perfect matching with delays for two sources. Theoret. Comput. Sci. 754, 122–129.

Gale, D., Shapley, L.S., 1962. College admissions and stability of marriage. Amer. Math. Monthly 69, 9–15.

Gurvich, I., Ward, A., 2015. On the dynamic control of matching queues. Stoch. Syst. 4 (2), 479–523.

Hardy, G.H., 1910. Orders of Infinity. Cambridge University Press.

Hizak, J., Logozar, R., 2011. A derivation of the mean absolute distance in one-dimensional random walk. Tehn. Glas. 5 (1), 10–16.

Janson, S., 1999. One, two and three times log n/n for paths in a complete graph with random weight. Probab. Comput. 8 (4), 347–361.

Jensen, J.L.W.V., 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. 30 (1), 175–193.

Kac, M., 1947. Random walk and the theory of Brownian motion. Amer. Math. Monthly 54, 369–391.

Kanoria, Y., Saban, D., Sethuraman, J., 2018. Convergence of the core in assignment markets. Oper. Res. 66 (3), 620–636.

Karp, R.M., Vazirani, U.V., Vazirani, V.V., 1990. An optimal algoritm for on-line bipartite matching. In: Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing. pp. 352–358.

Kerimov, S., Ashlagi, I., Gurvich, I., 2023. On the optimality of greedy policies in dynamic matching. Oper. Res..

Khintchine, A., 1924. Ü einen satz der wahrscheinlichkeitsrechnung. Fund. Math. 6 (1), 9–20.

Koenig, D., 1931. Grafok es matrixok. Mat. Fiz. Lapok 38, 116–119.

Krokhmal, P.A., Pardalos, P.M., 2009. Random assignment problems. European J. Oper. Res. 194, 1–17.

Kurino, M., 2014. House allocation with overlapping generations. Am. Econ. J. Microecon. 6 (1), 258–289.

Leshno, J.D., 2022. Dynamic matching in overloaded waiting lists. Amer. Econ. Rev. 112 (12), 3876–3910.

Linusson, S., Waestlund, J., 2004. A proof of Parisi's conjecture on the random assginment problem. Probab. Theory Related Fields 128, 419–440.

Loertscher, S., Muir, E.V., Taylor, P.G., 2022. Optimal market thickness. J. Econom. Theory 105383.

Lucas, R., Prescott, E., 1974. Equilibrium search and unemployment. J. Econom. Theory 7, 188–209.

McCall, J.J., 1970. Economics of information and job search. Q. J. Econ. 84 (1), 113–126.

Mehta, A., 2013. Online matching and ad allocation. Found. Trends Theor. Comput. Sci. 8, 265–368.

Mezard, M., Parisi, G., 1987. On the solution of the random link matching problems. J. Physique 48, 1451–1459.

Mortensen, D.T., 1970. A theory of wage and employment dynamics. In: Microeconomic Foundations of Employment and Inflation Theory. WW Norton, New York, pp. 167–211.

Peters, C.A.F., 1856. Ueber die Bestimmung des wahrscheinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem arithmetischen Mittel. Astron. Nachr. 44, 29–32.

Pinedo, M.L., 2012. Scheduling, vol. 29, Springer.

Potts, C.N., Kovalyov, M.Y., 2000. Scheduling with batching: A review. European J. Oper. Res. 120 (2), 228–249.

Roth, A.E., Xing, X., 1994. Jumping the gun: Imperfections and institutions related to the timing of market transactions. Amer. Econ. Rev. 84, 992–1044.

Roth, A.E., Xing, X., 1997. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. J. Polit. Econ. 105 (2), 284–329.

Shimer, R., 2007. Mismatch. Amer. Econ. Rev. 97, 1074–1101.

Su, X., Zenios, S.A., 2004. Patient choice in kidney allocation: The role of the queueing discipline. Manuf. Serv. Oper. Manag. 6 (4), 280–301.

Su, X., Zenios, S.A., 2005. Patient choice in kidney allocation: A sequential stochastic assignment model. Oper. Res. 53 (3), 443–455.

Su, X., Zenios, S.A., 2006. Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. Manage. Sci. 52 (11), 1647–1660.

Tonelli, L., 1909. Sull'integrazione per parti. Rend. Acc. Naz. Lincei 5, 246–253.

Ünver, M.U., 2010. Dynamic kidney exchange. Rev. Econom. Stud. 77, 372–414.

Waestlund, J., 2005. A proof of a conjecture of buck, chan, and robbins on the expected value of the minimum assignment. Random Struct. Algorithms 26, 237–251.

Waestlund, J., 2009. An easy proof of the $\zeta(2)$ limit in the random assignment game problem. Electron. Commun. Probab. 14, 261–269.

Wald, A., 1944. On cumulative sums of random variables. Ann. Math. Stat. 15, 283–296.

Walkup, D.W., 1979. On the expected value of a random assignment problem. SIAM J. Comput. 8 (3), 440–442.

Zenios, S.A., 2002. Optimal control of a paired-kidney exchange program. Manage. Sci. 48 (3), 328–342.