Interview Hoarding*

Vikram Manjunath

University of Ottawa

vikramma@gmail.com

Thayer Morrill

North Carolina State University

thayer_morrill@ncsu.edu

February 22, 2021

Abstract

Many centralized matching markets are preceded by interviews between the participants. We study the impact on the final match of an increase to the number of interviews one side of the market can participate in. Our motivation is the match between residents and hospitals where, due to the COVID-19 pandemic, interviews for the 2020-21 season of the NRMP match have switched to a virtual format. This has drastically reduced the cost to applicants of accepting interview offers. However, the reduction in cost is not symmetric since applicants, not programs, bore most of the costs of in-person interviews. We show that if doctors are willing to accept more interviews but the hospitals do not increase the number of interviews they offer, no doctor will be better off and potentially many doctors will be harmed. This adverse consequence results from a mechanism we describe as *interview hoarding.* We prove this analytically and characterize optimal mitigation strategies for special cases. We use simulations to extend the insights from our analytical results to more general settings.

Keywords: NRMP, Deferred acceptance, Interviews, Hoarding

1 Introduction

Perhaps the most well known application of matching theory is the entry-level labor market for physicians. In 2020, 37,256 positions were matched through the Na-

---

1

tional Resident Matching Program (NRMP). The matching process consists of two steps. First, each physician interviews with a set of residency programs. Second, programs and physicians submit rank-order lists—of only those they interview—to a centralized clearinghouse. This clearinghouse, run by the NRMP, matches physicians to residency programs using a version of Gale and Shapley (1962)'s Deferred Acceptance algorithm (Roth and Peranson, 1999).

In practice, both programs and applicants are constrained in the number of interviews they can take part in. Prior to the COVID-19 pandemic of 2020-21, interviews were done in person. These interviews were particularly costly for physicians since they not only had to bear travel expenses but had to use scarce vacation days. The cost to programs was mainly in terms of the time it takes. For the 2020-21 matching season, interviews were conducted virtually. While this dramatically decreased the cost of interviews for physicians, it did not change the costs much for the programs. We are interested in the implications of this change on the eventual match.

We focus our study particularly on the effects of doctors accepting more interview invitations without a corresponding increase to the number of invitations extended by programs. If some doctors accept more interviews and if the total number of invitations extended does not change, then some of the doctors necessarily receive fewer invitations. It seems natural to intuit that at least doctors with more interviews benefit from the lower costs even if those with fewer interviews are harmed. However, we show a surprising result: No physician is better off when more interviews can be accepted than under the previous arrangement. We prove this for a starting arrangement where the final matching is stable. We view stability as an equilibrium concept that describes a steady state of a market.

The intuition for this result is as follows. Consider a highly sought after physician; one who is offered interviews at the leading programs and who ultimately is matched with her favorite program. When interviews becomes cheaper, she will accept more interviews. However, as she would already have matched with her favorite program, the interviews she accepts are from inferior programs. These interviews do not help her: she ultimately matches with the same program as before. The interviews are, in effect, wasted. We refer to this as *interview hoarding*. Interview hoarding has a cascading affect. The physicians who otherwise would have filled these wasted interview slots now interview with programs they consider inferior. These physicians may have more interviews, but they do not have better interviews in a precise sense: every new interview a doctor has, she rates worse than the program she matched with before. Physicians are ultimately divided into three categories: physicians who hoard interview worse than their

eventual match; physicians who receive more but worse interviews; and physicians who receive fewer and worse interviews. The first category is indifferent between the new costs and the old. The latter two categories are harmed under the new cost. Thus, when physicians accept more interviews but programs do not react, the ultimate match is Pareto inferior from the physicians' perspective.[1]

Having shown that increases to doctors' abilities to accept interviews has adverse consequences, we turn to mitigation policies. We consider policies that limit the numbers of interviews that programs can offer and candidates can accept. Though there are, essentially, no such policies that *always* (for every preference profile) yield a stable final matching (Proposition 1), we characterize such policies for "common preferences" (Proposition 2). These are salient preference profiles where every doctor ranks the programs the same way and every program ranks the doctors the same way. The policies we characterize are such that there is a common cap on the number of interviews any program can offer or candidate can accept. We also show that if the programs interview capacities are fixed, say at l, then the number of blocking pairs increases and the match rate decreases as the doctors' interview cap gets further away from l (Proposition 3).

Our analytical results are suggestive of policies for more general settings where preferences are not quite common, but have a common component. We use simulations to show that the lessons from our analytical results hold up under weaker assumptions. Though the optimal cap on doctors' interview capacities depends on the parameters of the model—and in practice would have to be determined empirically—our simulations indicate that it is no higher than the number of interviews that the programs offer.[2]

Clearly, the main purpose of interviews is information acquisition. Through this channel, having more interviews can benefit a physician. For example, a physician benefits if by participating in more interviews she learns that a program whose interview she previously would have rejected is actually her favorite program. However, our aim is to isolate the effect of congestion in the interview step. We therefore assume that all agents have perfect information regarding their preferences.We show that an increase in this congestion can have an adverse impact on performance of the match itself unless either the residency programs or the NRMP reacts.

---

[1]The motivating example in Section 1.2 demonstrates that there need not exist a Pareto ranking from the programs' perspective.

[2]This is true whether we define optimality of a policy as maximizing the average proportion of positions that are filled or minimizing the average number of blocking pairs.

1.1 Related Literature

While there is a large literature on the post-interview NRMP match,[3] there are relatively few papers that incorporate the pre-match interview process. One of the first to explicitly model interviews in the classic one-to-one matching model is Lee and Schwarz (2017). In their model, before participating in a centralized, two-sided match, firms learn their preferences over workers by first engaging in costly interviews. They show that even if firms and workers interview with exactly the same number of agents, the extent of unemployment in the final match depends critically on the overlap between the sets of workers firms interview. Three other recent papers that incorporate pre-match interviews are Kadam (2015), Beyhaghi (2019), and Echenique et al. (2020).

Like us, Kadam (2015) considers the implications of loosened interview constraints for doctors. However, the focus is on strategic allocating scarce interview slots. For the sake of tractability, the analysis is for a stylized model of large markets. Under the assumption of common preferences, he shows that increasing student capacities may increase total surplus, but not in a Pareto-improving way. Moreover, match rate decreases. He also highlights that when preferences are not necessarily common, the effect is ambiguous since increased interview capacities dilutes doctors' signaling ability.

Beyhaghi (2019) also performs a strategic analysis of a stylized large market model. However, she considers a slightly different set up with *application* caps for doctors and interview caps for programs. While similar, application caps are not exactly the same as interview caps: they constrain the number of programs a doctor can express interest in at the outset of the interview matching phase, but not the number of interviews she can accept at the end. In her model, inequity in the application caps decreases expected total surplus. Moreover, when interview capacity is low, low application caps are socially desirable.

In our model, the agents do not choose interviews strategically. Determining the optimal set of interviews is closely related to the portfolio choice problem considered by Chade and Smith (2006). They solve for the optimal portfolio when an agent chooses a portfolio of costly, stochastic options, but ultimately may only accept one of the options. If one were to apply the optimal solution to the interview scheduling problem, one would have to pin down precisely the probability of any given pair matching. This is what makes strategic analysis of the problem intractable without severe simplifying assumptions (such as those in the papers we have mentioned above).

---

[3]See the multitude of papers following Roth and Peranson (1999).

The analysis of Echenique et al. (2020) is methodologically closest to ours. They explain a puzzling empirical pattern resulting from the NRMP match: 46.3% of the physicians were matched to their favorite residency programs and 71.1% where matched to one of their favorite three programs. These statistics seem to contradict surveys indicating that many doctors have similar preferences over residency programs. They provide an explanation for this phenomenon by pointing out the importance of the interviewing process that precedes the match. Roughly speaking, the pre-match interviewing process truncates the preferences that the physicians submit to the actual NRMP clearinghouse. Therefore, a proper interpretation is not that the physicians matched with their most preferred programs but rather that they matched with their most preferred programs *among those they interviewed with*.

Our work is complementary with these works in the sense that they highlight the importance of understanding the prematch interviews to properly evaluating the NRMP match itself.

1.2 Motivating Example

We present the intuition behind the welfare loss from increased interview capacity for doctors with a simple example. Consider a market with four doctors $\{d_1,...,d_4\}$ and four hospitals $\{h_1,...,h_4\}$. The agents' preferences are as follows:

d_1	d_2	d_3	d_4	h_1	h_2	h_3	h_4
h_1	h_2	h_2	h_1	d_1	d_1	d_2	d_1
h_2	h_3	h_1	h_2	d_2	d_2	d_1	d_4
h_4	h_1	h_3	h_3	d_3	d_4	d_3	d_3
h_3	h_4	h_4	h_4	d_4	d_3	d_4	d_2

Suppose that the interview capacities of the doctors and hospitals are:

d_1	d_2	d_3	d_4	h_1	h_2	h_3	h_4
1	2	1	1	1	1	2	2

Interviews are initially offered by hospitals: h_1 invites d_1, h_2 invites d_1, h_3 invites d_2 and d_1, and h_4 invites d_1 and d_4. As d_1 can accept only one invitation, she rejects h_2, h_3, and h_4. Doctors d_2 and d_4 do not reject their invitations from h_3 and h_4, respectively. Hospitals h_2, h_3, and h_4 then offer interviews to d_1, d_3, and d_3, respectively. Doctor d_3 rejects h_4's invitation. After h_4 invites and is rejected by d_2, the final interviews are:

d_1	d_2	d_3	d_4
h_1	$\{h_2, h_3\}$	h_3	h_4

The final matching is computed by applying the doctor-proposing Deferred Acceptance algorithm to the agent preferences (restricted to agents they interview with). The outcome is therefore:

d_1	d_2	d_3	d_4
h_1	h_2	h_3	h_4

The interview schedule is well functioning in the sense that the ultimate outcome is stable with regards to the actual (as opposed to restricted) preferences.

Now suppose doctor 1 is able to accept an additional interview (and all other interview capacities remain the same). Doctor 1 now accepts h_2's invitation. Doctor 2 eventually accepts hospital 4's invitation, and the interview schedule is:

d_1	d_2	d_3	d_4
$\{h_1, h_2\}$	$\{h_3, h_4\}$	h_3	h_4

This leads to the final matching:

d_1	d_2	d_3	d_4
h_1	h_3		h_4

Doctor d_1 does not benefit from the additional interview. Since the original matching was stable, the interview she adds is with a hospital that she finds worse than her original match. However, her acceptance of hospital 2's invitation comes at the expense of both doctors 2 and 3: both now receive worse assignments. In fact, the final matching is no longer stable. That none of the doctors are better off is not special to this example—we show that this is generally true (Theorem 1). The programs, however, are not unanimously better or worse off: h_3 is better off while h_2 is worse off.

2 The Model

A **market** consists of a triple $(\mathbf{D}, \mathbf{H}, \mathbf{P})$, where: D is a finite set of **doctors**; H is a finite set of **hospitals**; and P is a profile of strict **preferences** for the doctors and hospitals. We assume that there are at least two doctors and two hospitals: $|D| \geq 2$ and $|H| \geq 2$. For each $h \in H$, $\mathcal{P}_h$ is the set of strict preferences over $D \cup \{h\}$,

and for each $d \in D$, $\mathcal{P}_d$ is the set of strict preferences over $H \cup \{d\}$. The set of preference profiles is $\mathcal{P} \equiv \times_{i \in H \cup D} \mathcal{P}_i$.

Like Echenique et al. (2020), we combine an interview phase with a matching phase. The former involves many-to-many matching while the latter is a standard one-to-one matching problem (Roth and Sotomayor, 1990). A matching is a function $\mu : H \cup D \to H \cup D$ such that $\mu(h) \in D \cup \{h\}$, $\mu(d) \in H \cup \{d\}$, and $\mu(d) = h$ if and only if $\mu(h) = d$. We say that (d, h) are a **blocking pair** to matching μ if $h\, P_d\, \mu(d)$ and $d\, P_h\, \mu(h)$. A matching is **stable** if it does not have a blocking pair.

A *many-to-many* matching is a function $\nu : H \cup D \to 2^{H \cup D}$ such that $\nu(d) \subseteq H$, $\nu(h) \subseteq D$, and $h \in \nu(d)$ if and only if $d \in \nu(h)$.

For each $h \in H$, let $\iota_h \in \mathbb{N}$ be **h's interview capacity**. Similarly, for each $d \in D$, let $\kappa_d \in \mathbb{N}$ be **d's interview capacity**. We call the profile $(\iota, \kappa) = ((\iota_h)_{h \in H} (\kappa_d)_{d \in D})$ the **interview arrangement**. An **interview matching** is a many-to-many matching ν such that for every doctor d, $|\nu(d)| \leq \kappa_d$ and for every hospital h, $|\nu(h)| \leq \iota_h$.

An interview matching ν is **pairwise stable** if there is no doctor-hospital pair (d, h) such that $h \notin \nu(d)$ but:

- either $|\nu(h)| < \iota_h$ or there exists a $d' \in \nu(h)$ such that $d\, P_h\, d'$, and

- either $|\nu(d)| < \kappa_d$ or there exists a $h' \in \nu(d)$ such that $h\, P_d\, h'$.

Two-step process: Given (ι, κ), we call the final matching, which is the culmination of the following two step process, the **(ι, κ)-matching** . In the first step, where hospitals interview doctors, we ignore the informational effects and focus solely on congestion. Given $P \in \mathcal{P}$:

Step 1: Interview matching ν is the hospital-optimal many-to-many stable matching where the capacities of the hospitals and doctors are given by ι and κ respectively. This can be computed by applying the hospital-proposing DA: each $h \in H$ is matched with up to ι_h doctors and each $d \in D$ is matched with up to κ_d hospitals. Since we ignore the informational aspect of the problem, the input to DA is a choice function for each agent that is responsive to her preference relation and constrained by her interview capacity.[4] Hospital-proposing DA is an approximation of the decentralized process by which hospitals invite doctors in rounds, extending invitations to further doctors when invitations are declined.

---

[4]For the sake of completeness, we define in Appendix A the acceptant and responsive choice functions that we appeal to while running DA to compute the interview matching.

Step 2: The (ι, κ)-matching is chosen by doctor-proposing DA. The input to DA is the true preference profile restricted to the interview match, $(P_i|_{\nu(i)})_{i \in D \cup H}$.

Given $P \in \mathcal{P}$, we say that (ι, κ) **is adequate** if the (ι, κ)-matching at P is stable. If κ is adequate at each $P \in \mathcal{P}$, then we say that it is **globally adequate**. We interpret (ι, κ) being adequate at a profile P as a sign that the market is functioning well. Otherwise, a blocking pair could alter their behavior to improve their lot. In other words, using stability as our notion of equilibrium, (ι, κ) being adequate is equivalent to the market being in equilibrium.

Finally, we define a welfare comparison between matchings. Given a pair of matchings μ and μ', we say that **no doctor prefers** μ' **to** μ if, for each $d \in D$, $\mu(d) \, R_d \, \mu'(d)$.

3 Welfare Impact of Increased Interviews

Our aim is to study how a change in the cost of interviewing impacts a market. The starting point is a market that is at equilibrium. Starting from such an equilibrium, the goal is to understand the welfare consequences of a shock that permits doctors to accept more interviews. That is, starting with $P \in \mathcal{P}$ and (ι, κ) that is adequate at P, we consider an increase to the doctors' interview capacities to κ' and compare the (ι, κ)-matching to the (ι, κ')-matching.

The doctor who accepted more interviews in our example in Section 1.2 did not benefit from it. Our main result shows that this is true in general.

Theorem 1. *Starting at an adequate arrangement, doctors do not benefit from increases to their interview capacities. That is, if (ι, κ) is adequate at P and κ' is such that, for each $d \in D$, $\kappa'_d \geq \kappa_d$, then the no doctor prefers the (ι, κ')-matching to the (ι, κ)-matching.*

Proof. Let ν and μ be the interview and final matchings respectively, under (ι, κ). Similarly, let ν' and μ' be the interview and final matchings under (ι, κ'). We frame the temporal language below in reference to a hypothetical change in doctors' interview capacities from κ ("before") to κ' ("after").

We first establish a number of properties of the outcome from the interview step. The intuition for these results comes from one of the classical results in two-sided matching theory: When the set of men increases, no man benefits from this increased competition while no woman is harmed.[5] In our setting, an increase in

---

[5]See Theorem 2.25 of Roth and Sotomayor (1990).

the number of interviews a doctor can participate in plays the role of additional men participating in the market.

Lemma 1. *No doctor rejects a hospital it previously interviewed with.*

Proof. Suppose not. In the interview matching step (under capacities κ'), let d be the first doctor to reject a hospital h that she interviewed with under capacities κ. As d has at least as much interview capacity, she must have received a new proposal from some hospital h'. As h' did not propose to d before, it must have been rejected by some doctor $d' \in \nu(h)$, a doctor it previously interviewed. But this contradicts d being the first doctor to reject a hospital it previously interviewed with. □

We cannot say whether a doctor prefers her interviews under κ versus κ' as we only have a doctor's preferences over individual hospitals and not sets of hospitals. However, we show—in a specific sense—that while a doctor may get new interviews, she does not get better interviews.

Lemma 2. *No doctor has a new interview better than her previous matching: if $h \in \nu'(d) \setminus \nu(d)$, then $\mu(d) P_d h$.*

Proof. Suppose not. Let d be the first doctor when DA is run during the interview step under capacities κ' to receive a proposal from a hospital $h \notin \nu(d)$ such that $h P_d \mu(d)$. As h did not previously propose to d, h must have been rejected by a doctor that it previously interviewed. This contradicts Lemma 1. □

In the classical result, no man benefits from the increased competition due to additional men and also no woman is harmed. An analogous result holds in our framework. A hospital either has the same set of interviews; additional interviews; or she interviews new doctors who she prefers to her previous interviews. In any scenario, the hospital's set of interviews (weakly) improves.

Lemma 3. *Suppose a hospital h interviews a doctor d under κ'. If h previously interviewed d' and prefers d' to d, then h continues to interview d': if $d \in \nu'(h)$, $d' \in \nu(h)$, and $d' P_h d$, then $d' \in \nu'(h)$.*

Proof. As $d' P_h d$, h proposes to d' before it proposes to d when DA is run in the interview step under κ'. By Lemma 1, h is not rejected by any doctor it previously interviewed. As h proposes to d under κ', it must have already proposed but not be rejected by d'. Therefore, h continues to interview d'. □

To complete the proof of Theorem 1, we show that if a doctor is rejected by a hospital during the matching step under κ, then she is not matched to that hospital under κ'. We proceed by induction on the round (of DA in the interview step under κ) in which the doctor was rejected, and our inductive hypothesis is that if doctor d was rejected by hospital h in round k under κ, then under κ', either she no longer interviews with h or she is rejected in round k or earlier.

For the base step, consider a doctor d that was rejected by hospital h in the first round under κ, and let d' be the doctor h tentatively accepts. If d no longer interviews with h ($d \notin \nu'(h)$), then we are done. Therefore, suppose $d \in \nu'(h)$. By Lemma 3, since h prefers d' to d and it interviews d, it also interviews d' ($d' \in \nu'(h)$). Doctor d' does not have any new interviews with a hospital it prefers to h since $h\,R_{d'}\,\mu(d')$ and by Lemma 2 she does not get a new interview with a hospital she prefers to $\mu(d')$. Therefore, d' continues to propose to h in the first round even under the new capacities and d continues to be rejected by h in favor of d' or possibly a doctor h prefers even more.

To complete the inductive argument, suppose that doctor d was rejected by hospital h in favor of doctor d' in round k under κ. If $h \notin \nu'(d)$, then we are done. Otherwise, again by Lemma 3, $d' \in \nu'(h)$. Under κ, d' proposes to h in round k or earlier. Therefore, d' was rejected by all hospitals she interviewed with and prefers to h in an earlier round. By the inductive hypothesis, for any hospital h' that rejected d' under κ, either d' no longer interviews with h' or h' has already rejected d' by round k under κ'. Therefore, under κ', either d' proposes to h in round k or in a previous round. In either case, by round k, under κ', h has already received a proposal it prefers to d. Therefore, doctor d will be rejected by hospital h under κ' in round k or earlier.

This shows that if d was rejected by hospital h under the old capacities, then d is not matched to h under the new capacities. Note that d has no new interviews with a hospital she prefers to $\mu(d)$. Therefore, if $h\,P_d\,\mu(d)$ and $h \in \nu'(d)$, then $h \in \nu(d)$ and h rejected d in some round under the old capacities. Therefore, h also rejects d under the new capacities. In particular, under the new capacities, d is not matched to a hospital she prefers to $\mu(d)$. $\qquad\square$

Theorem 1 tells us that doctors increasing the number of interviews they accept will either have no impact on the resulting matching or will make the new matching Pareto worse from the doctors' perspective. The example in Section 1.2 illustrates that there are instances where increasing the interview capacity does result in a Pareto inferior outcome. Note that this example is not pathological. Lemmas 1 and 2 demonstrate the root cause of the inferior match. Doctors at the "top" of the market—those that are highly sought after—add interviews but every

10

new interview they accept is with a hospital that is worse than the one they will eventually be assigned to. This is what we refer to as *interview hoarding*. A poor final matching occurs when a hospital's interviews are all accepted by doctors who will eventually reject it.

Our conclusion from Theorem 1 is that the impact of virtual interviews for the 2020-21 season of the NRMP, *if the residency programs do not react* sufficiently, will be a lower match rate. We expect the most highly sought after physicians to have more interviews than usual while less demanded physicians will have few or no interviews. It is essential that residency programs either increase the number of interviews they conduct or rethink their strategies for offering interviews.

4 Adequate Arrangements

Theorem 1 assumes that the initial profile of interviews was adequate in the sense that the outcome of the two-step process is a stable matching. We interpret this assumption as a characteristic of a well-functioning market in steady state equilibrium. A natural question is how many interviews need to take place and what does the distribution of interviews need to be in order for an interview profile to be adequate. Of course, in general, the answer will depend on characteristics of the market such as the ratio of doctors to hospitals and how correlated or aligned preferences are. However, we are able to provide tight characterizations for certain "end-point" cases which provide intuition for more general markets.

4.1 Globally Adequate Arrangements

In studying adequate arrangements, we first ask about worst case performance: what arrangements are adequate *for every* preference profile? It turns out that only very extreme arrangements satisfy this property. We characterize these arrangements in our next result.

Proposition 1. *Arrangement* (ι, κ) *is globally adequate, if and only if either*

1. *every doctor and every hospital has only unit interview capacity—that is, for each* $d \in D, \kappa_d = 1$ *and for each* $h \in H, \iota_h = 1$—*or*

2. *every doctor and every hospital has high interview capacity—that is, for each* $d \in D, \kappa_d \geq \min\{|D|, |H|\}$ *and for each* $h \in H, \iota_h \geq \min\{|D|, |H|\}$.

Proof. We first prove necessity. Suppose that (ι, κ) is globally adequate.

We start by establishing that if one doctor or hospital has greater than unit interview capacity, then every doctor and hospital has interview capacity of at least two. Stated differently, if any doctor or hospital has unit capacity, then all doctors and hospitals have unit capacity. We denote by ν the interview matching and by μ the (ι, κ)-matching.

Claim 1. *1. If there is $d \in D$ such that $\kappa_d > 1$, then for each $d' \in D, \kappa_d \geq 2$ and for each $h \in H, \iota_h \geq 2$, and*

2. If there is $h \in H$ such that $\iota_h > 1$, then for each $h' \in H, \iota_h \geq 2$ and for each $d \in D, \kappa_d \geq 2$.

Proof. We prove only the first statement as the proof of the second statement is analogous—it requires only a reversal of the roles of doctors and hospitals.

Suppose, for the sake of contradiction, there is a globally adequate (ι, κ) where there exists a $d_1 \in D$ such that $\kappa_{d_1} > 1$ and a $h_2 \in H$ such that $\iota_{h_2} = 1$. Let $h_1 \in H \setminus \{h_2\}$ and $d_2 \in D \setminus \{d_1\}$. Consider $P \in \mathcal{P}$ where each doctor ranks h_1 first and h_2 second, and each hospital ranks d_1 first and d_2 second. All hospitals offer an interview to d_1 and as $\kappa_{d_1} > 1$, d_1 accepts interviews from at least h_1 and h_2. Since $h_2 = 1$, h_2 only interviews d_1. Let μ be the (ι, κ)-matching. Since (ι, κ) is adequate, μ is stable, so $\mu(d_1) = h_1$, as h_1 and d_1 are mutual favorites. Therefore, $\mu(h_2) = h_2$ as h_2 only interviews d_1. Note that (d_2, h_2) forms a blocking pair of μ as $h_2 P_{d_2} \mu(d_2)$, since $\mu(d_2) \notin \{h_1, h_2\}$, and $d_2 P_{h_2} h_2$. This contradicts the stability of μ and thus the assumption that (ι, κ) is globally adequate. We have therefore established that if there is $d \in D$ such that $\kappa_d > 1$, then for each $h \in H$, $\iota_h \geq 2$.

We now prove that if there $d_1 \in D$ such that $\kappa_{d_1} > 1$, then for each $d \in D$, $\kappa_d \geq 2$. Suppose for the sake of contradiction, that there is $d_2 \in D$ such that $\kappa_{d_2} = 1$. Let $h_1, h_2 \in H$. Consider $P \in \mathcal{P}$ such that each doctor ranks h_1 first and h_2 second, and each hospital ranks d_1 first and d_2 second. As we have shown above, $\iota_{h_1}, \iota_{h_2} \geq 2$, so both h_1 and h_2 offer interviews to both d_1 and d_2. Since h_1 is her favorite hospital, d_2 accepts its offer. Thus, $\nu(d_2) = \{h_1\}$. However, $\mu(d_1) = h_1$ since d_1 and h_1 are mutual favorites, so $\mu(d_2) = d_2$. This means that (d_2, h_2) form a blocking pair of μ as the only hospital d_2 prefers to h_2 is h_1. This contradicts the stability of μ and thus the assumption that (ι, κ) is globally adequate. □

We complete the proof of necessity by showing neither a doctor nor a hospital can have an intermediate capacity.

Claim 2. *There is no $d \in D$ such that $1 < \kappa_d < \min\{|D|, |H|\}$, and there is no hospital h such that $1 < \iota_h < \min\{|D|, |H|\}$.*

Proof. We prove this statement for the case where $|D| \leq |H|$. The proof when $|H| < |D|$ is symmetric.

Suppose for contradiction that $d_1 \in D$ is such that $\kappa_{d_1} = k$ where $1 < k < |D|$. Let $P \in \mathcal{P}$ be such that for i from 1 through $k + 1$:

$$P_{d_1} : h_2, h_3, \ldots, h_{k+1}, h_1, \ldots$$
$$P_{h_i} : h_i, h_1, \ldots, h_{i-1}, h_{i+1}, \ldots$$
$$P_{h_1} : d_1, d_2, \ldots$$
$$P_{h_i} : d_i, d_1, \ldots, d_{i-1}, d_{i+1}, \ldots$$

We have constructed the preference profile P such that:

- For each i from 1 through $k + 1$, d_i and h_i are matched in every stable matching.

- Each of the $k + 1$ hospitals $h_1, \ldots, h_{k+1}$ offers d_1 an interview.

- Doctor d_1 accepts interview offers from hospitals $h_2, \ldots, h_{k+1}$, but not from h_1.

The first and third points are immediate consequences of the preferences. The second is a consequence of the first part of Claim 1: since $\kappa_{d_1} > 1$, every hospital has interview capacity of at least two and d_2 is its second favorite doctor. However, this contradicts the definition of μ as the (ι, κ)-matching, since $h_1 \notin \nu(d_1)$ yet by stability, $h_1 = \mu(d_1)$.

A similar construction shows that there is no $h \in H$ such that $1 < \iota_h < |D|$. Suppose for contradiction that $h_1 \in H$ is such that $\iota_{h_1} = l$ where $1 < l < |D|$. Let $P \in \mathcal{P}$ be such that for i from 1 through $l + 1$:

$$P_{d_1} : h_1, h_2, \ldots$$
$$P_{d_i} : h_i, h_1, \ldots, h_{i-1}, h_{i+1}, \ldots$$
$$P_{h_1} : d_2, d_3, \ldots, d_{l+1}, d_1$$
$$P_{h_i} : d_i, d_1, \ldots, d_{i-1}, d_{i+1}, \ldots$$

By the second part of Claim 1, since $\iota_{h_1} > 1$, every doctor has capacity of at least two. Therefore:

- For each i from 1 through $l + 1$, d_i and h_i are matched in every stable matching.

- Each of the l doctors $d_2, \ldots, d_{k+1}$ accepts an interview from h_1 an interview.

13

- Hospital h_1 does not offer d_1 an interview.

Thus, $h_1 \notin \nu(d_1)$, so $h_1 \neq \mu(d_1)$. This contradicts the stability of μ, the (ι, κ)-matching, and in turn the assumption that (ι, κ) is globally adequate. □

We now turn to sufficiency. If every agent has an interview capacity of one, then the interview matching is actually a matching. Moreover, it is a stable matching. So, suppose that each agent has an interview capacity of at least $\min\{|D|, |H|\}$. If $|D| = |H|$, then the interview matching involves an interview between every mutually acceptable doctor-hospital pair. This means that the (ι, κ)-matching is the doctor optimal stable matching under the unrestricted preferences, which is stable. We now show, that even if $|D| < |H|$ or $|D| > |H|$, the (ι, κ)-matching, μ, is stable. Suppose the doctor-hospital pair (d, h) blocks μ. By defintion of μ as the (ι, κ)-matching, if $h P_d \mu(d)$ and $d P_h \mu(h)$, then $h \notin \nu(d)$.

Suppose $|D| < |H|$. Since $\iota_h \geq |D|$, h would have offered an interview to d and have been rejected during the interview matching step, so $\nu(d)$ contains κ_d hospitals that d prefers to h. Since $h P_d \mu(d)$, and $\mu(d) \in \nu(d) \cup \{d\}$, this means $\mu(d) = d$. Then, d is rejected by every hospital in $\nu(d)$ during the application of DA in the second step. However, $|\nu(d)| = \kappa_d \geq |D|$ and since d is acceptable to every hospital in $\nu(d)$, she is only rejected when another doctor applies. However, this implies that when DA terminates in the second step, every hospital in $\nu(d)$ has tentatively accepted some doctor other than d, which is a contradiction—there are not enough such doctors.

Suppose $|H| < |D|$. Since $\kappa_d \geq |H|$, d does not reject any interviews she is offered. Since $h \notin \nu(d)$, h offers interviews to and has them accepted by $\iota_h \geq |H|$ doctors whom it prefers to d. Since $d P_h \mu(h)$, h does not receive a proposal from any $d' \in \nu(h)$ during the application of DA in the second step since it finds all such d' better than d. This implies that each $d' \in \nu(h)$ is tentatively accepted by some hospital other than h when DA terminates, which is a contradiction—there are not enough such hospitals. □

Proposition 1 highlights a previously overlooked role that the interview step plays in determining whether or not the ultimate NRMP match is stable. While interviews are necessary for agents to gain information, we learn from Proposition 1 that interviews can also act as a bottleneck. Even with complete information, once any agent is capable of participating in more than one interview, all agents must interview with essentially the entire market to be certain that the ultimate match is stable.

4.2 Homogeneous Arrangements

The distribution of interviews is an essential factor in the stability of the NRMP match. So, it is natural to consider market interventions when the interview step is out of balance. Our motivating question is what happens when there is an increase to the number of interviews doctors can accept. The most straightforward intervention is to cap the number of interviews an agent can participate it. Here we consider a **homogeneous arrangement**: all doctors face the same cap and all hospitals face the same cap, but we allow the doctor and hospital caps to potentially differ. In other words, the arrangement would be described by two numbers: an interview capacity $l \in \mathbb{N}$ for hospitals and an interview capacity $k \in \mathbb{N}$ for doctors. The pair (l, k) corresponds to the arrangement (ι, κ) where for each $h \in H, \iota_h = l$ and for each $d \in D, \kappa_d = k$.

By Proposition 1 a homogenous arrangement (l, k) can only be globally adequate if $l = k = 1$ or $l, k \geq \min\{|D|, |H|\}$. Nonetheless, (l, k) *may* be adequate for a specific profile of preferences. One might ask whether, starting at a profile $P \in \mathcal{P}$ and arrangement (l, k) that is adequate at P, if the comparative statics with respect to l and k are consistent. The following examples demonstrate that this is not so. It may be that, depending on P, increasing k by 1 renders a previously adequate arrangement inadequate, or the opposite. In other words, the effect of the increase to k is specific to P. Similarly for l.

Example 1. *Either incrementing or decrementing l or k can render an adequate arrangement inadequate.*
Suppose $|D| = |H| = 3$ and consider $P \in \mathcal{P}$ such that for each $i = 1, 2, 3,$[6]

P_{h_i}	P_{d_i}
d_1	h_1
d_2	h_2
d_3	h_3
h_i	d_i

For P, $(2, 2)$ is adequate: the interview matching is ν such that $\nu(h_1) = \nu(h_2) = \{d_1, d_2\}$ and $\nu(h_3) = \{d_3\}$. So, the (l, k)-matching is μ such that for each $i = 1, 2, 3$, $\mu(h_i) = d_i$, which is the unique stable matching.

We now observe that if we increment or decrement either l or k by one, the arrangement is no longer adequate for P. In other words, none of $(1, 2), (3, 2), (2, 1),$

---

[6]This can be embedded into a larger problem instance.

and $(2,3)$ is adequate for P. We summarize the interview matching and the (l,k)-matching for each of these below.

(l,k)	interview matching	(l,k)-matching
$(1,2)$	$v(h_1) = v(h_2) = \{d_1\}, v(h_3) = \{d_2\}$	$\mu(h_1) = d_1, \mu(h_3) = d_2, \mu(h_2) = h_2, \mu(d_3) = d_3$
$(3,2)$	$v(h_1) = v(h_2) = D, v(h_3) = \{\}$	$\mu(h_1) = d_l, \mu(h_2) = d_2, \mu(h_3) = h_3, \mu(d_3) = d_3$
$(2,1)$	$v(h_1) = \{d_1, d_2\}, v(h_2) = \{d_3\}, v(h_3) = \{\}$	$\mu(h_1) = d_1, \mu(h_2) = d_3, \mu(h_3) = h_3, \mu(d_2) = d_2$
$(2,3)$	$v(h_1) = v(h_2) = v(h_3) = \{d_1, d_2\}$	$\mu(h_1) = d_1, \mu(h_2) = d_2, \mu(h_3) = h_3, \mu(d_3) = d_3$

All four of the final matchings are unstable.   ∘

The mechanics of Example 1 are robust and it is not by accident that $(2,2)$ is adequate to start with. The preferences in the example have a particular salient configuration, which we focus on here. A profile $P \in \mathcal{P}$ has **common preferences** if all doctors rank the hospitals in the same way, and all hospitals rank the doctors in the same way. To further restrict the definition, we also require that each doctor finds each hospital acceptable and each hospital finds each doctor acceptable. That is, for each pair $d, d' \in D$ and each pair $h, h' \in H$, $P_d|_H = P_{d'}|_H$, $P_h|_D = P_{h'}|_D$, $d\,P_h\,h$, and $h\,P_d\,d$.[7]

As we see from Example 1, a result like Proposition 1 does not hold if we restrict ourselves to common preferences. Our next result is a characterization of homogeneous arrangements that are adequate for common preferences.[8]

Proposition 2. *Under common preferences, a homogeneous arrangement (l,k) is adequate if and only if $l = k$ or $l, k \geq \min\{|D|, |H|\}$.*

Proof. Let $P \in \mathcal{P}$ be such that there are common preferences. Let $\{d_t\}_{t=1}^{|D|}$ and $\{h_t\}_{t=1}^{|H|}$ be enumerations of D and H respectively such that every hospital prefers d_t to d_{t+1} and every doctor prefers h_t to h_{t+1}. Let $m = \min\{|D|, |H|\}$. There is a unique stable matching μ^*, such that for each $t = 1, \dots, m$, $\mu^*(h_t) = d_t$.

Let v be the interview matching under (l,k) and μ be the (l,k)-matching.

First, we show that (l,k) is adequate for P only if $l = k$ or $l, k \geq \min\{|D|, |H|\}$. Suppose $l \neq k$. If $l < k$ and $l < \min\{|D|, |H|\}$, then for each $t = 1, \dots, k$, $v(h_t) = \{d_1, \dots, d_l\}$. In particular, $d_k \notin v(h_k)$ so $\mu(h_k) \neq d_k$. On the other hand, if $l > k$ and $k < \min\{|D|, |H|\}$, then for each $t = 1, \dots, l$, $v(d_t) = \{h_1, \dots, h_k\}$. In particular,

---

[7]Under common preferences there is, obviously, a unique stable matching.

[8]The characterization of Proposition 2 does not hold for arrangements that are not homogeneous. For a counterexample, suppose $|D| = 4$, $|H| = 3$, there is $d \in D$ such that $\kappa_d = 3$, for each $d' \in D \setminus \{d\}, \kappa_{d'} = 2$, there is $h \in H$ such that $\iota_h = 4$, and for each $h' \in H \setminus \{h\}, \iota_{h'} = 2$. For any common preferences, (ι, κ) is adequate.

16

$h_l \notin v(d_l)$ so $\mu(d_l) \neq h_l$. In either case, the (l,k)-matching is not stable so (l,k) is not adequate.

Now, we show that if $l = k \leq m$, then (l,k) is adequate. For each $t = 1,\dots,m$, let $\underline{t} = \lfloor \frac{t-1}{l} \rfloor$. Then, for each $t = 1,\dots,i$, $v(h_t) = \{d_{\underline{t}+1},\dots,d_{\underline{t}+l}\}$ and $v(d_t) = \{d_{\underline{t}+1},\dots,d_{\underline{t}+l}\}$. Thus, for each $t = 1,\dots,m$, $\mu(h_t) = d_t$. So (l,k) is adequate at P.

Finally, if $l,k \geq m$, then for each $t = 1,\dots,m$, $v(d_t) \supseteq \{h_1,\dots,h_m\}$. Since $h_t \in v(d_t)$, $h_t = \mu(d_t)$. So (l,k) is adequate at P. $\qquad\square$

If the hospitals' interview capacity is fixed at some specific l, an important policy decision is where to set the doctors' interview cap, k. Proposition 2 says that the optimal value for k is exactly at l whether the objective is to minimize the number of blocking pairs or to maximize the match rate (the proportion of positions that are filled). Our next result sheds light on this objective.

Proposition 3. *Fix the hospitals' interview capacity at l and consider k and k' such that either $k' < k \leq l$ or $l \leq k < k'$. Suppose $P \in \mathcal{P}$ has common preferences. The (l,k')-matching has more blocking pairs and a lower match rate than the (l,k)-matching.*

Proof. Let $P \in \mathcal{P}$ be such that there are common preferences. Let $\{d_t\}_{t=1}^{|D|}$ and $\{h_t\}_{t=1}^{|H|}$ be enumerations of D and H respectively such that every hospital prefers d_t to d_{t+1} and every doctor prefers h_t to h_{t+1}.

Let $m = \min\left\{\left\lfloor \frac{|H|}{k} \right\rfloor, \left\lfloor \frac{|D|}{l} \right\rfloor\right\}$. The interview matching is such that for each d_t, if $t \leq ml$,

$$v(d_t) = \{h_{(n-1)k+1},\dots,h_{nk}\} \text{ where } n \text{ is such that } (1-n)l < t \leq nl$$

if $ml < t \leq (m+1)l$,

$$v(d_t) = \begin{cases} \{h_{mk+1},\dots,h_n\} & \text{if } |H| \geq mk+1 \\ \emptyset & \text{otherwise} \end{cases} \quad \text{where } n = \min\{|H|,(m+1)k\}$$

and if $(m+1)n < t$, $v(d_t) = \emptyset$.

We first consider the case where when $k > l$ and show that the number of matched hospitals is decreasing in k and that the number of blocking pairs is increasing in k.

Given P and its restriction to v, the (l,k)-matching, μ, at P is such that for each d_t, if $t \leq ml$,

$$\mu(d_t) = h_{(n-1)k+(t \bmod l)}, \text{ where } n \text{ is such that } (n-1)l < t \leq nl,$$

17

if $ml < t \leq (m+1)l$,

$$\mu(d_t) = \begin{cases} h_{mk+(t \bmod l)} & \text{if } |H| \geq mk + (t \bmod l) \\ d_t & \text{otherwise,} \end{cases}$$

and if $(m+1)l < t$, $\mu(d_t) = d_t$.

Let $n = \min\{|H| - mk, |D| - ml\}$. Given the (l,k)-matching above, the set of matched hospitals is

$$\{h_{ik+s} : i = 0,\dots,m-1, s = 1\dots,l\} \cup \{h_t : t = mk+1,\dots,mk+n\}.$$

Therefore, the number of matched hospitals is $ml + n$. Holding l fixed, this is decreasing in k.

The (l,k)-matching, is blocked by all pairs consisting of an unmatched hospital and any doctor with a higher index. That is, $(h_t, d_{t'})$ such that $t \leq mk$, $t-1 \bmod k \geq l$ and $t' > t$. These are the only pairs the block it. Thus, the number of blocking pairs is

$$\sum_{n=0}^{m-1} \sum_{i=l+1}^{k} |D| - (nk+i).$$

Holding l fixed, this is increasing in k.

Now, we consider the case where $k < l$ and show that the number of matched hospitals is increasing in k and the number of blocking pairs is decreasing in k.

Given P and its restriction to ν, the (l,k)-matching at P is such that for each h_t, if $t \leq mk$,

$$\mu(h_t) = d_{(n-1)l+(t \bmod k)}, \text{ where } n \text{ is such that } (n-1)l < t \leq nl,$$

if $mk < t \leq (m+1)k$,

$$\mu(h_t) = \begin{cases} h_{ml+(t \bmod k)} & \text{if } |D| \geq ml + (t \bmod k) \\ h_t & \text{otherwise} \end{cases}$$

and if $(m+1)k < t$, $\mu(h_t) = h_t$.

Let $n = \min\{|H| - mk, |D| - ml\}$. Given (l,k)-matching above, the set of matched hospitals is $\{h_t : t \leq mk + n$. Therefore, the number of matched hospitals is $mk + n$. Since $k < l$, this is weakly increasing in k.

The (l,k)-matching, is blocked by all pairs consisting of an unmatched doctor and any hospital with a higher index. That is, $(d_t, h_{t'})$ such that $t \leq ml$, $t-1 \bmod l \geq$

18

k and $t' > t$. These are the only pairs the block it. Thus, the number of blocking pairs is

$$\sum_{n=0}^{m-1} \sum_{i=k+1}^{l} |D| - (nl + i).$$

Holding l fixed, this is decreasing in k. □

5 Simulations

Our analytical results are of two sorts. On one hand, Theorem 1 applies without restrictions on preferences. However, it only has something to say about doctors' welfare and only in regards to perturbations to an equilibrium arrangement. On the other hand, when we focus on common preferences, Proposition 2 and Proposition 3 deliver a clearcut policy prescription. In this section, we use simulations to bridge the gap. This allows us to consider how changes in the doctors' interview capacities affect hospitals' welfare, match rates, stability, and so on, in a more general setting.

While there is evidence that preferences do indeed have a common component (Agarwal, 2015; Rees-Jones, 2018), agents care about "fit" as well. Moreover, an idiosyncratic component is to be expected. We adopt the random utility model of Ashlagi et al. (2017).[9] Each hospital $h \in H$ has a common component to its quality, x_h^C and a "fit" component, x_h^F. Similarly, each doctor $d \in D$ has a common component to her quality, x_d^C and a fit component, x_d^F. The utilities that h and d enjoy from being matched to one another are

$$u_h(d) = \beta x_d^C - \gamma \left(x_h^F - x_d^F \right)^2 + \varepsilon_{hd}$$
$$\text{and}$$
$$u_d(h) = \beta x_h^C - \gamma \left(x_h^F - x_d^F \right)^2 + \varepsilon_{dh}$$

respectively, where ε_{hd} and ε_{dh} are drawn independently from the standard logistic distribution. Each x_h^C, x_h^F, x_d^C, and x_d^F is drawn independently from the uniform distribution over $[0, 1]$. The coefficients β and γ weight the common and fit components respectively. When β and γ are both zero, preferences are drawn uniformly at random. As $\beta \to \infty$, these approach common preferences. As γ increases, preferences become more "aligned": the fit, which is orthogonal to the common component, becomes more important.

---

[9]This, in turn, is adapted from Hitsch et al. (2010).
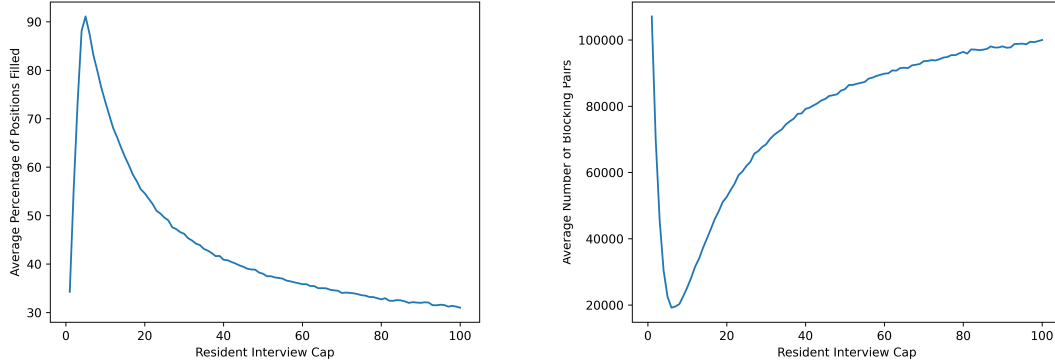
(a) The average match rate is highest at $k = 5$ (91.095 pairs).

(b) The average number of blocking pairs is lowest at $k = 6$ (19,183.43 pairs).

Figure 1: We vary k from 1 to 100 with l fixed at 25.

Our simulated market has 400 hospitals.[10] We have chosen the number of doctors to be 470.[11] The parameters for the random utility model are $\beta = 40$ and $\gamma = 20$. Since our interest is in the effects of changes to doctors' interview capacities, we fix hospital interview capacities at $l = 25$.

Our first simulation results involve varying k from 1 to 100.[12] Figure 1a shows that the match rate increases and then decreases. On the other hand, Figure 1b shows that the number of blocking pairs decreases and then increase. These results are consistent with what we learn from Proposition 3. However, since preferences are not common, the match rate does not reach 100% and the number of blocking pairs remains positive even at the optimal k—that is, the arrangement is not adequate. Moreover, the optimal k does not equal to l.

Our next set of results evaluate a hypothetical policy of restricting doctors to a maximum of 5 interviews. We choose this as a candidate policy as it is optimal in

---

[10]The NRMP match is broken down into smaller matches by specialty. In 2020, among 50 specialties for PGY-1 programs, the largest had 8,697 positions, 10th largest had 849 positions, the 25th largest had 38 positions, the 49th largest had one position and the smallest had no positions. This data is available from the NRMP. Our chosen number of hospitals is comparable to the 70th percentile among specialties.

[11]There were, on average, 0.85 PGY-1 positions per applicant in the 2020. Our chosen number of students reflects this ratio.

[12]We have chosen this upper bound to be high enough that further increases have little effect. Thus, we interpret this as doctors being essentially unconstrained in how many interviews they can accept.

terms of maximizing the match rate (Figure 1a). We compare this policy with the benchmark of no intervention where doctors are completely unconstrained.

Figure 2a shows the distribution of the number of doctors who prefer their match under the optimal k over the benchmark as well as the distribution of those with the opposite preference. We see that the former is considerably far to the right of the latter. Though Theorem 1 applies only to when the starting arrangement is adequate, Figure 2a shows that the lesson from that result does extend beyond. Figure 2b shows the same distributions, except for hospitals. Despite the fact that Theorem 1 does not address hospitals' welfare, our simulations show that more hospitals prefer the optimal k than leaving the doctors unconstrained. The policy also has the benefit of drastically decreasing the number of blocking pairs. Figure 2c shows the distribution of excess blocking pairs when we compare the matching under $k = 5$ to the benchmark matching where doctors are unconstrained. Finally, we compare the distribution of interviews among the doctors between the two arrangements in Figure 2d. The constraint limiting doctors to $k = 5$ interviews binds for many doctors. An implication of this is that significantly more doctors receive zero interviews when doctors are unconstrained. This is consistent with the intuition that if interviews were costless for doctors, then highly sought after doctors would hoard interviews and others would be left with nothing.

We next compare the proposed intervention ($k = 5$) to a hypothetical "ideal" world where there is no interview stage and the final match is based on actual preferences. Figure 3a displays the distributions of the numbers of doctors who prefer the match under the intervention as well as the numbers with the opposite preference. Figure 3b displays the analogous distributions for hospitals. Though the hospitals are typically worse off, the match under interview constraints and the intervention is *not Pareto dominated* by the "ideal" match: typically, many doctors are *better off*.

Finally, we consider the possibility that the NRMP could set not only a cap on interviews that doctors can accept, but can also control the number of interviews that hospitals offer. From Proposition 2, we know that if preferences are common, the match rate would be maximized where $l = k$. Figure 4 shows that, even when preferences are not exactly common, there are still optimal combinations of l and k, but they typically involve $l > k$.
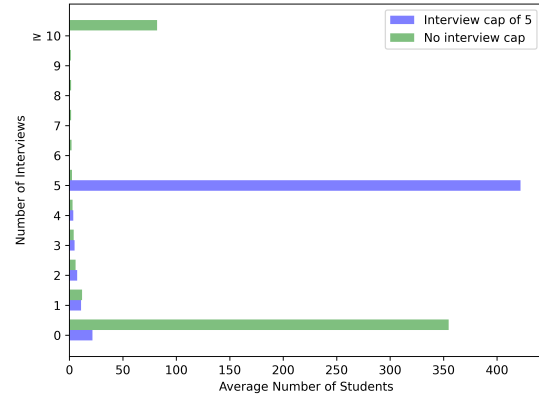
21
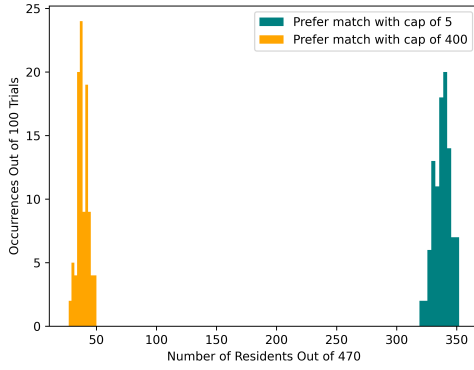
(a) Distribution of the number of doctors who prefer their match at $k = 5$ over being unconstrained and vice versa.
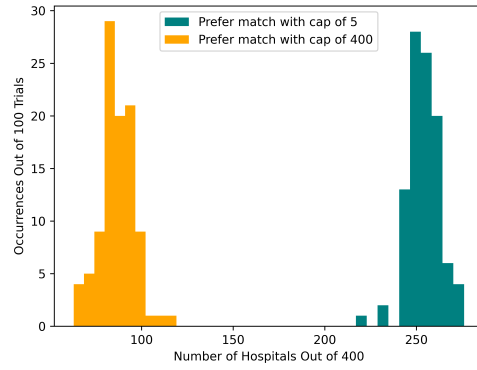
(b) Distribution of the number of hospitals who prefer their match at $k = 5$ over the doctors being unconstrained and vice versa.
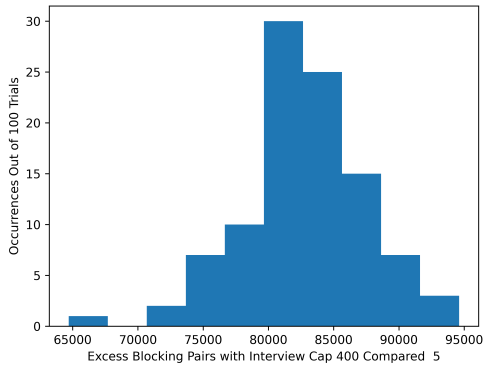
(c) Distribution of the number of excess blocking pairs when doctors are unconstrained over the number of such pairs at $k = 5$.

(d) Distribution of interviews at $k = 5$ without a cap. The uncapped distribution vanishes with the number of interviews reaching the hundreds.

Figure 2: Comparisons of the intervention of capping doctors' interview capacities at $k = 5$ to leaving doctors unconstrained.
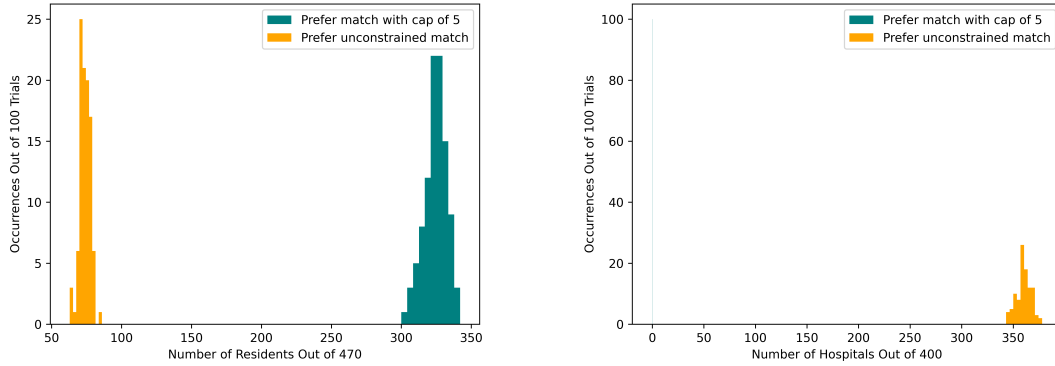
22

(a) Distribution of the number of doctors who prefer their match at $k = 5$ over their match without an interview stage and vice versa.

(b) Distribution of the number of hospitals who prefer their match at $k = 5$ over their match without an interview stage and vice versa.

Figure 3: Comparisons of welfare under the intervention $k = 5$ to the hypothetical scenario where there is no interview step.

6 Conclusion

The 2020-21 global pandemic has had a significant impact on the way interviews are conducted. We anticipate that it will also impact the number of interviews doctors participate in. Through our simulations and theoretical results, we predict that unless hospitals also increase the number of interviews they offer, the 2021 NRMP match will result in a lower percentage of positions being filled and a less stable matching.

In future years, the NRMP should consider policies to mitigate these effects. Our analysis supports the idea of interview caps and our simulations provide evidence that such a policy would reduce the bottleneck created by the interview step. Such caps can be implemented with very limited centralization, for instance, using a ticket system.

Even if such interventions are not possible in the very short run, our policy prescription is that residency programs should be advised to increase the number of candidates they interview relative to previous years.

Design of a fully centralized clearinghouse, is an area that remains open. As earlier work on the interview pre-markets have shown, strategic analysis has only been tractable under very stringent assumptions (Kadam, 2015; Lee and Schwarz,
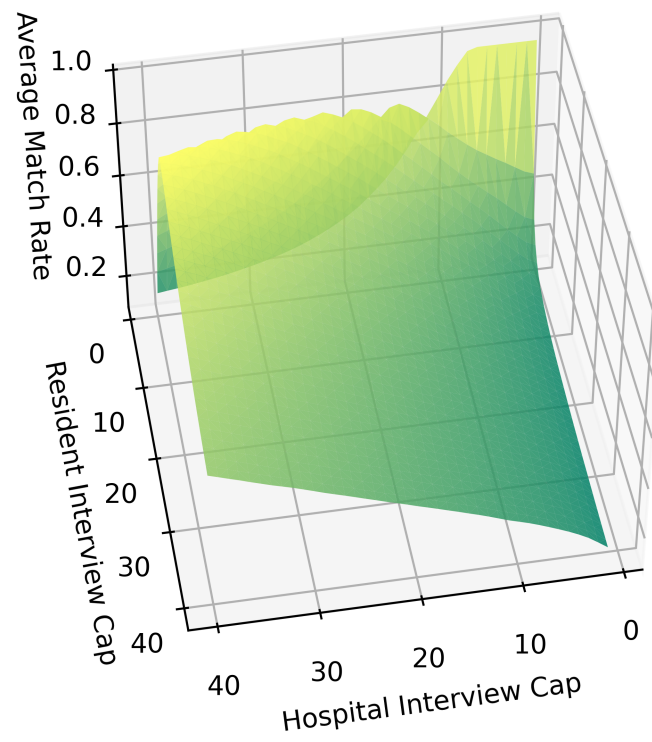
23

Figure 4: Match rate as a function of l and k

2017; Beyhaghi, 2019). Nonetheless, the current paper adds to the evidence (along with Echenique et al. (2020)) that a more holistic approach that includes the interview stage is critical.

This interview driven bottleneck is likely a factor other matching contexts as well, including fully decentralized labor markets. For example, we expect it to affect the junior market for economists. This labor market typically consists of short interviews followed by on-campus visits. Physical constraints typically limit both the number of short interviews and on-campus visits a candidate is able to accept. With virtual interviews and virtual "fly-outs", we expect candidates to accept more of both than they would otherwise. As a result, we could expect the same bottleneck in the economics job market as in the NRMP match.

References

Agarwal, Nikhil (2015) "An Empirical Model of the Medical Match," *American Economic Review*, Vol. 105, No. 7, pp. 1939–78, July. 19

Ashlagi, Itai, Yash Kanoria, and Jacob D. Leshno (2017) "Unbalanced Random Matching Markets: The Stark Effect of Competition," *Journal of Political Economy*, Vol. 125, No. 1, pp. 69–98. 19

Beyhaghi, Hedyeh (2019) "Approximately-optimal Mechanisms in Auction Design, Search Theory, and Matching Markets," Ph.D. dissertation, Cornell University. Chapter 5: Two-Sided Matching with Limited Number of Interviews. 4, 25

Chade, Hector and Lones Smith (2006) "Simultaneous Search," *Econometrica*, Vol. 74, No. 5, pp. 1293–1307. 4

Echenique, Federico, Ruy Gonzalez, Alistair Wilson, and Leeat Yariv (2020) "Top of the Batch: Interviews and the Match." 4, 5, 7, 25

Gale, David and Lloyd Shapley (1962) "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, Vol. 69, pp. 9–15. 2

Hitsch, Gunter J., Ali Hortaçsu, and Dan Ariely (2010) "Matching and Sorting in Online Dating," *American Economic Review*, Vol. 100, No. 1, pp. 130–63, March. 19

Kadam, Sangram V (2015) "Interviewing in Matching Markets." 4, 23

Lee, Robin S. and Michael Schwarz (2017) "Interviewing in two-sided matching markets," *The RAND Journal of Economics*, Vol. 48, No. 3, pp. 835–855. 4, 23

Rees-Jones, Alex (2018) "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match," *Games and Economic Behavior*, Vol. 108, pp. 317–330. 19

Roth, Alvin E. and Elliott Peranson (1999) "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, Vol. 89, No. 4, pp. 748–780, September. 2, 4

Roth, Alvin E. and Marilda A. Oliveira Sotomayor (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monographs: Cambridge University Press. 7, 8

Appendices

A   Choice Functions for Interview Step

In the interview step, we compute a many-to-many matching. However, each doctor and each hospital only ultimately matches to at most one other partner, and each has strict preferences over partners. This necessitates the definition of a choice function over sets of partners. For our analytical results, we focus on acceptant choice functions that are responsive to preferences over partners and constrained by interview capacity. That is, given $P \in \mathcal{P}$,

- From the set $H' \subseteq H$, each $d \in D$ chooses the κ_d best elements of H' according to P_d:

$$
C_d(H') = \begin{cases} \{h \in H' : h\, P_d\, d\} & \text{if } |\{h \in H' : h\, P_d\, d\}| \leq \kappa_d \text{ and} \\[2ex] B \subseteq \{h \in H' : h\, P_d\, d\} & \text{such that } |B| = \kappa_d \text{ and for each } h \in B \\ & \text{and each } h' \in H' \setminus B, h\, P_d\, h' \text{ otherwise.} \end{cases}
$$

- From the set $D' \subseteq D$, each $h \in H$ chooses the κ_h best elements of D' accord-

ing to P_h:

$$C_h(D') = \begin{cases} \{d \in D' : d\,P_h\,h\} & \text{if } |\{d \in D' : d\,P_h\,h\}| \leq \iota_h \text{ and} \\\\ B \subseteq \{d \in D' : d\,P_h\,h\} & \text{such that } |B| = \iota_h \text{ and for each } d \in B \\ & \text{and each } d' \in D' \setminus B, d\,P_h\,d' \text{ otherwise.} \end{cases}$$