HERDING WITH HETEROGENEOUS ABILITY: AN APPLICATION TO ORGAN TRANSPLANTATION

By

Stephanie De Mel, Kaivan Munshi, Soenje Reiche

and Hamid Sabourian

September 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2308R1

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Herding with Heterogeneous Ability: An Application to Organ Transplantation*

Stephanie De Mel† Kaivan Munshi‡ Soenje Reiche§ Hamid Sabourian¶

September 23, 2024

Abstract

There are many economic environments in which an object is offered sequentially to prospective buyers. It is often observed that once the object for sale is turned down by one or more agents, those that follow do the same. One explanation for this phenomenon is that agents making choices further down the line rationally ignore their own assessment of the object and herd behind their predecessors. Our research extends the canonical herding model by allowing agents to differ in their ability to assess the quality of the offered object. We develop novel tests of herding based on this ability heterogeneity and also examine its efficiency consequences, applied to organ transplantation in the U.K. We find that herding is common but that the information lost due to herding does not substantially increase false discards of good organs or false acceptances of bad organs. Our counter-factual analysis indicates that this is due (in part) to the high degree of heterogeneity in ability across transplant centers. In other settings, such as the U.S., where organ transplantation is organized very differently and the ability distribution will not be the same, the inefficiencies due to herding might well be substantial.

Keywords. Social Learning. Herd behavior. Organ Transplantation. Agent Heterogeneity.
JEL. D82. D83. I18.

*We are grateful to Kourosh Saeb-Parsy and Patrick Trotter for their collaboration on this project and to numerous seminar participants for their constructive comments. Research support from the Keynes Fund, Cambridge-INET, and the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010 is gratefully acknowledged. We are responsible for any errors that may remain.

†Westminster City Council sdemel6@gmail.com
‡Yale University and Toulouse School of Economics kaivan.munshi@yale.edu
§Yale University soenje.reiche@yale.edu
¶University of Cambridge hs102@cam.ac.uk

1 Introduction

There are many economic environments in which prospective buyers, acting sequentially, must choose whether or not to acquire an object. Examples of such environments include venture capital and property development, where a startup or a piece of land is offered for sale, and the labor market, notably the draft in professional sports leagues and the academic job market. It is often observed in these environments that once the object for sale is turned down by one or more agents, those that follow do the same. One explanation for this correlation in decisions is that the object is independently assessed to be of poor quality by all agents. An alternative explanation, which goes back to seminal contributions by Banerjee (1992) and Bikhchandani et al. (1992) is that agents who must make choices further down the line (rationally) ignore their own assessment of the object's quality and herd behind their predecessors. Our research extends the canonical herding model by allowing agents to differ in their ability to assess the quality of the offered object. We develop novel tests of herding based on this ability heterogeneity and also examine its consequences for efficiency.

The setting for our research is the transplantation of deceased donor organs in the United Kingdom. The organ transplant program in the U.K. is organized around a nationwide network of centers (hospitals). When a deceased donor organ becomes available, all patients on the National Transplant Registry are assigned a priority rank based on a predetermined allocation algorithm. Transplant centers are offered the organ in order of their patients' priority, until the organ either is accepted or, having deteriorated with time, is no longer viable and is discarded by the National Health Service Blood and Transplantation (NHSBT). We will see that the organization of the U.K. transplant program makes it an ideal setting for our tests of herding. In particular, most decisions are made at early positions (on average, the third position) and, hence, patient-organ mismatch and organ deterioration, which increase by position, will be shown to be less relevant. This is also a setting in which information-based herding, associated with uncertainty about organ quality, may have large practical consequences. Currently, demand outstrips the availability of both livers and kidneys, the two organs that dominate transplantation activity in the U.K. and that constitute the focus of our analysis. A natural question to ask is whether herding, and the accompanying discarding of good organs, has contributed to this shortfall.

The example that follows explains why transplant centers might rationally follow the decisions of the centers that preceded them, and why such herding could generate inefficiencies. When a center is offered an organ, it can either accept or reject. Suppose that there are two types of organs: good (G) organs, which should be accepted and bad (B) organs, which should be rejected. Each transplant center makes an assessment of the quality of the organ that is made available to it. This assessment, which we characterize as an information signal, is not directly observed by other transplant centers. As is common in the literature on herd behaviour; e.g. Bikhchandani et al. (1992), Anderson and Holt (1997), and other references cited in Chamley (2004), we assume that signals are binary: good (g) and bad (b). These signals are informative; i.e. centers are more likely to receive a g (b) signal when an

organ is good (bad). Centers have a common prior about the quality of offered organs and a cut-off belief above which they will accept an organ. Moreover, the common prior and the cut-off belief are such that a center always follows its signal in first position; this is simply saying that the posterior belief following a g (b) signal lies above (below) the cut-off belief. The next center in line for a given organ is only offered that organ if the first center rejects, which implies that it must have received a b signal. This shifts the second center's prior below the common prior. If this downward shift is sufficiently large, then the second center will reject even if it receives a g signal (it will certainly reject if it receives a b signal). Although it may be individually rational for centers to ignore their signals and herd behind their predecessors in this fashion, useful information contained in these signals is lost to all centers that follow and can result in the under-utilization of viable organs.[1]

Over the past three decades, the literature on information-based herding has advanced on multiple fronts. Early theoretical contributions examined the robustness of information cascades to alternative signal and choice structures (see Gale (1996) and Chamley (2004) for overviews). Theoretical work on financial markets; e.g. Avery and Zemsky (1998) and Park and Sabourian (2011) showed that herding can arise even when the cost of conformity (the asset price) is increasing in the number of agents who have made the same (investment) decision. In parallel, empirical contributions in finance and development sought to establish the key implication of social learning, a general phenomenon subsuming herding but not necessarily involving information loss, which is that agents condition their decisions on their predecessors' decisions (Lakonishok et al., 1992; Grinblatt et al., 1995; Wermers, 1999; Foster and Rosenzweig, 1995; Munshi, 2004; Conley and Udry, 2010). More recently, the theoretical focus has shifted to model misspecification (Bohren, 2016; Frick et al., 2019) and non-Bayesian learning on networks (Golub and Jackson, 2010; Mobius et al., 2015; Banerjee et al., 2021). The empirical literature has developed sharper tests of social learning by exploiting experimentally induced variation in predecessors' decisions (Dupas, 2014) and has moved to estimating structural models of herding (Zhang, 2010; Cipriani and Guarino, 2014). Our analysis advances the empirical herding literature by incorporating a new dimension of agent heterogeneity; the ability to assess the quality of the object under consideration.

Our analysis proceeds in three steps. First, we derive tests of herding with heterogeneous ability. When choices are sequential, the most straightforward test of herding, and social learning more generally, is that the same center is more likely to reject an organ when it is further down the line because it has observed more rejections. However, other explanations for this observation are also available. For example, lower quality organs, which are less likely to be accepted at any position on average, travel further down the line even when centers make decisions independently.[2] Furthermore,

[1] In general, herding arises when agents do not fully incorporate their signals in their decisions. An "information cascade" is an extreme type of herding in which agents completely ignore their own signals when they follow their predecessors and, as a result, agents further down the line learn nothing from their decisions. In our model, there are two actions, accept or reject, two types of organs (states), G and B, and two signals, g and b. Centers either follow their signals or ignore them completely and, hence, herding and information cascades are synonymous.

[2] This negative selection is conceptually related to the well known dynamic selection bias that arises when individuals with heterogeneous ability make decisions sequentially and drop out non-randomly from the sample over time; e.g.

patient priority is determined, in part, by the organ-patient match. Patients (and their associated centers) lower in line will have a worse match on average and, moreover, the organ is more likely to have deteriorated with time. It is possible to control for these factors, as in Zhang (2010), but such conditioning is always imperfect. We avoid this problem by developing tests of herding and its associated information loss that leverage variation in decisions at the *same* position. This variation arises because (i) centers differ in their ability to distinguish between good and bad organs, and (ii) the ordering of centers varies from one organ to the next depending on the priority of their patients (and not on center ability).

We implement the tests of herding with administrative data obtained from NHSBT. This data, described in Section 2, covers the universe of deceased-donor livers and kidneys offered between 2006 and 2015 in the United Kingdom. It includes the sequence of centers that were offered each organ, as well as their decisions, which – with the possible exception of the final center in every sequence – must necessarily be rejections. Although there is no variation in predecessors' decisions, centers differ in their ability to distinguish between good and bad organs. In our model, introduced in Section 3, higher ability centers are better at detecting both types of organs. They thus pass forward lower quality organs on average; i.e. a greater fraction of bad organs, which implies that the centers that follow are more likely to reject. We build on this insight to construct a measure of center ability in Section 4.1 and to develop tests of herding in the sub-sections that follow.

Our tests of herding are based on the idea that centers always follow their signals in first position, in the absence of any other information, but can herd and ignore their signals in second position, as in the example above. The first test in Section 4.2 restricts attention to decisions in second position, focusing on the interaction of center 1 ability, q_1, and center 2 ability, q_2. When center 2 follows its own signal, it is more responsive to the deterioration in the organ pool that results from center 1 being of relatively high ability when it too is of higher ability. That is, we expect the effect of an increase in $q_1 \cdot q_2$ on the rejection decision in second position to be positive. When center 2 ignores its own signal and herds behind its predecessor, however, this interaction effect could be reversed. This is because lower ability centers in second position are more likely to abandon their signals and reject with certainty, especially when following higher-ability centers. A notable feature of this test of herding is that we are able to derive empirically verifiable conditions under which the $q_1 \cdot q_2$ effect is negative (positive) in the presence of herding. In contrast, as noted above, this effect is always positive when herding is absent.

Our second test of herding in Section 4.3 restricts attention to decisions in third position. Suppose, to begin with, that all centers follow their own signals. As above, higher ability centers pass on a worse pool of organs to those that follow, and it can be shown that marginal increases in q_1 and q_2 have the same (positive) effect on center 3's rejection decision. However, once we introduce the possibility of herding, which can only happen in second position, the effect of an increase in q_2 is strictly smaller

Cameron and Heckman (1998).

than the effect of a corresponding increase in q_1. This is because centers that herd in second position, and reject with certainty, regardless of their signal, do not alter the quality of the organ pool and, thus, the rejection probability of center 3. A notable feature of this test is that we are able to verify its implications with a substantially restricted sample of organs for which q_1 and q_2 are approximately the same.

In our model, centers with heterogeneous ability make decisions based on their assessment of organ quality. In practice, these decisions will also depend on idiosyncratic organ-center specific factors such as the patient-organ match and organ deterioration. As discussed in Section 4.4, such factors will not bias our estimates of center ability nor will they confound the tests of herding. In addition, we examine the possibility that other dimensions of center heterogeneity that might determine decision-making could undermine our results. Variation in the quality of organs received by centers could potentially bias our estimates of center ability, which are based on the overall rejection rate in first position. As documented in Section 2, however, average organ quality at that position, based on "risk indices" described below, does not vary across centers. Heterogeneity in unobserved pickiness; i.e. the cut-off belief above which a center accepts, could also generate variation in the first-position rejection rate. However, this source of heterogeneity cannot explain observed decisions in second and in third position, in the absence of herding, as discussed in Section 4.4 and shown formally in the Appendix.

The second step in the analysis estimates and validates the model. While it is sufficient to construct a composite, center-specific, measure of ability, q, for the herding tests, we need center and organ-quality specific measures of ability to verify the assumptions of the model, to examine its goodness of fit with the data, and for the efficiency analysis. The challenge here is that the quality of a given organ in the data is not known to the researcher, even *ex post*, particularly if it is not accepted. We thus incorporate risk indices that have recently been developed in the transplantation literature (Watson et al., 2012; Collett et al., 2017) and that are negatively associated with the probability that an organ is good, to estimate the β parameter (the probability of receiving a b signal with a B organ) and the γ parameter (the probability of receiving a g signal with a G organ) for each center in Section 5.1. We establish that the β, γ parameters are positively associated with the composite q parameter used in the herding tests, cross-validating the estimated parameters and verifying the assumption in the model that more able centers are better at detecting both good and bad organs. The β, γ parameters are used, in addition, to estimate the remaining parameters of the model in Section 5.2: the fraction of good organs in the population, which serves as the prior belief for all centers in position 1, and the cut-off belief that an organ is good above which centers accept. The estimated fraction of good organs is substantially lower for livers than for kidneys. This finding independently validates the ability measures that we construct for the herding tests, separately by organ type, and is consistent with our interpretation of the test results.

We complete the validation of the model by examining its goodness of fit with the data in Section 5.3. The model's parameters are estimated at the first and second positions and, hence, we are

interested in seeing how well the model matches the data at higher positions. Despite its parsimonious structure, we find that the rejection rate predicted by the model matches closely with the data at each position. This contrasts with the performance of an alternative no learning model in which centers make decisions independently; rejection rates continue to increase across positions due to the negative selection, but predicted rejection rates are now systematically lower than the rates observed in the data. The goodness of fit results complement the tests of herding, providing independent evidence that centers are learning from their predecessors' (rejection) decisions. Moreover, given the parameter estimates and the sequence of centers associated with each organ, we can determine whether a given center at a given position is herding. We find that herding is common – this occurs for 48% of decisions for livers, and 17% of decisions for kidneys – which leads us to the final step in the analysis where we examine the efficiency consequences of herding and its associated information loss.

One advantage of specifying a binary signal structure with two types of organs is that we can incorporate a continuum of center abilities in the model. A second advantage is that we can conveniently measure efficiency against a first-best benchmark in which all good organs are accepted and all bad organs are discarded. Although the risk index provides an objective measure of organ quality, it does not indicate whether a given organ in our data is good (should be accepted) or bad (should be discarded). We thus consider hypothetical "good" organs and hypothetical "bad" organs in the efficiency analysis reported in Section 6. For each organ, we draw centers randomly from our pool of centers and then order them without regard to their ability, in line with the ordering scheme that is currently in place. We draw a signal for each selected center, based on the type of organ and the center's estimated ability, and then use the model to predict beliefs and accompanying decisions at each position. In addition to our herding model, we also predict outcomes with a counter-factual pooled information model where the information loss due to herding is eliminated; i.e. in which signals received by centers that reject regardless of their signal are made available to centers positioned later in line. The results of this exercise can be summarized as follows: (i) both false discard rates and false acceptance rates are very similar with the two models of learning, (ii) false discard rates and false acceptance rates are low overall (approximately 3 percent).

We complete the analysis by examining the contribution of ability heterogeneity to the preceding results. We do this by comparing false discard rates and false acceptance rates at different (counterfactual) levels of ability heterogeneity, holding constant average ability. The inefficiency generated by the pooled information model is relatively insensitive to ability heterogeneity. In contrast, we find that the inefficiency due to herding, relative to the pooled information model, is declining with ability heterogeneity. The false discard rate, in particular, would double if centers were homogeneous in their abilities. In general, the effect of a mean-preserving increase in ability heterogeneity on the prevalence of herding is ambiguous; herding will increase when higher ability centers go earlier in the sequence and decrease when they go later. In our data, the prevalence of herding, and ability conditional on herding, decline with heterogeneity. This explains why the information loss and accompanying

inefficiency due to herding is relatively low.

Our results are relevant for analyses of organ transplantation in other settings. Early contributions to the organ transplantation literature in economics; e.g. Roth et al. (2004), Roth et al. (2007), focussed on kidney exchange and its associated matching problem involving multiple (living) donor-patient pairs. More recently, economists have begun to examine the allocation of deceased donor organs. With these organs, uncertainty about objective organ quality is an important consideration (in addition to the donor-patient mismatch) in decision-making. The fact that risk indices of organ quality have been developed in the U.S. and the U.K., with the objective of ultimately informing center decision-making, suggests that the medical community is well aware of this consideration. Moreover, since centers are making decisions sequentially, uncertainty about organ quality is likely to result in herding, with its potential inefficiencies. There is no *a priori* reason why the conditions that reduce herding inefficiencies in the U.K. should hold in other settings, such as the U.S., where organ transplantation is organized very differently and where the ability distribution will not be the same. Nevertheless, recent contributions to the transplantation literature in economics; e.g. Agarwal et al. (2020) Agarwal et al. (2021), ignore herding in their (U.S. based) analyses. Deceased donor organs account for the majority of transplanted organs in the U.S. and the U.K. and this line of research is likely to grow over time. It thus seems especially important to incorporate relevant elements of center decision-making, including social learning, in future research.

2 Institutional Setting

The shortage of suitable donor organs has always been the primary challenge faced by organ transplant programs. In response to this challenge, many countries, including the United Kingdom, have established national allocation schemes for the distribution of organs supplied by deceased donors. Organs obtained from deceased donors are classified according to the manner of death as either DBD (donation after brain death) or DCD (donation after cardiac death). Although DBD and DCD organs do not vary systematically with respect to *ex ante* quality and the same broad allocation protocols are utilized by the National Health Service Blood and Transplant (NHSBT) for both types of organs, DCD organs are useable for a shorter period of time before they must be discarded from the donor pool and set aside for research (Watson and Dark, 2012).

Our analysis focuses on livers and kidneys, which dominate transplantation activity in the United Kingdom: NHSBT statistics indicate that over 80% of livers and kidneys obtained from DBD donors in 2014-2015 were transplanted, while the corresponding statistics for pancreases, hearts, and lungs were less than 35%. A national allocation scheme, in place since 2006 for kidneys and 2015 for livers, offers DBD organs to patients based on their priority. This priority is determined by the patient-organ match and patients' characteristics, including their age, time on the wait list, and the distance between the donor hospital and transplant center. Prior to 2015, DBD livers were first offered to patients on a super-urgent national list, then to local centers, and, finally, to other centers. With DCD organs,

transportation delays are substantially more costly. These organs are thus typically allocated to local centers before being offered nationally.

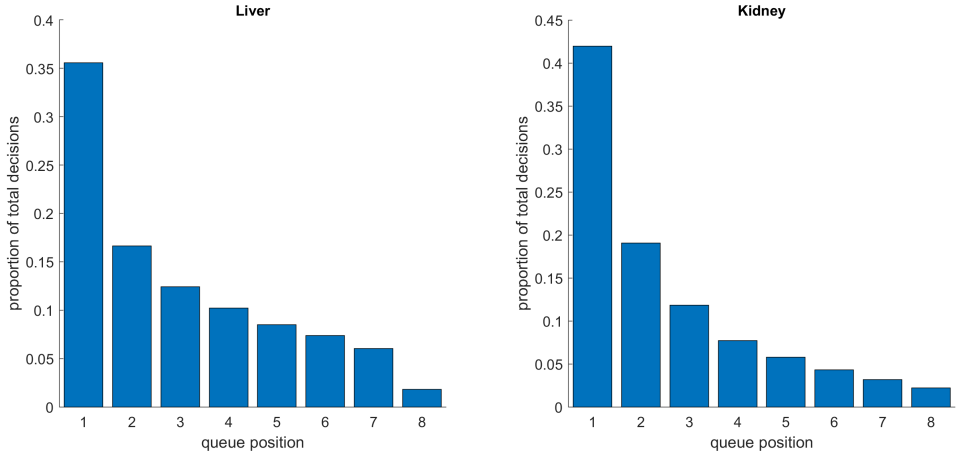
When an organ becomes available, it is offered to patients in order of their priority or to their centers, based on their proximity, as described above. Each patient's hospital (transplant center) has 45 minutes to accept or decline the offer. The attending surgeon has an open-ended conversation with the NHSBT administrator about the characteristics of the organ and the donor, as well as other factors that are relevant for that particular case, before arriving at a decision. Each center that is offered an organ can either accept or reject it. If an organ is rejected, it is offered to the next center in line, unless NHSBT assesses that the condition of the organ has deteriorated to the point that it is no longer useable, in which case it is discarded, i.e. set aside for research. There are thus two possible end-points for an organ: it is accepted or it is discarded. Prior to either end-point, all decisions must necessarily be rejections.

The deterioration that results in an organ being discarded can be caused by delays in retrieving the organ (warm ischaemia), which applies to DCD organs, or by subsequent delays in transplantation (cold ischaemia), which applies to both DBD and DCD organs. Based on NHSBT guidelines, DBD (DCD) livers should be transplanted within 12 (6) hours after retrieval, while the corresponding cutoffs for kidneys are 18 (12) hours. This is a narrow time window, leaving room for just a few centers to make decisions before an organ is discarded by NHSBT. The data that we use to test the model consists of the sequence of decisions taken by centers for each deceased-donor organ (liver or kidney) that was offered for transplantation in the 2006-2015 period.[3] Based on this data, as observed in Figure 1, 35% of observed decisions for livers are in first position, with a steep decline in the fraction of decisions at higher positions. For kidneys, over 40% of decisions are in first position, followed by an even steeper decline in the fraction of decisions at higher positions. There are relatively few decisions past the eighth position for either type of organ. The analysis that follows will thus be restricted to the first eight positions.

The allocation of deceased-donor organs in the United Kingdom differs in important respects from the allocation procedure in the United States. Zhang (2010), using data on kidney donations in Texas, documents that on average an organ is accepted by the 34^{th} patient in line, who has already turned down 15 offers. Such long queue lengths are possible because organs are only discarded after 48 hours. Under these circumstances, the condition of the organ becomes a major consideration in decision-making, particularly at higher positions. The mismatch between organ and recipient also becomes relevant (in Zhang's data, kidneys are accepted as late as the 77^{th} position). Given this mismatch, patients consider (and reject) many organs before finally accepting and, hence, dynamic considerations enter the decision rule. Zhang (2010), Agarwal et al. (2020), and Agarwal et al. (2021), who all study kidney allocation in the United States, model the acceptance decision as an optimal stopping problem.

[3] As discussed above, a new National Kidney Allocation Scheme was initiated in 2006 and a new National Liver Allocation Scheme was initiated in 2015. The analysis thus covers a period during which each type of organ was allocated in a uniform manner.

Figure 1: Distribution of Decisions by Queue Position



Note: Statistics based on all organs offered for transplantation, 2006-2015.

The institutional environment in the United Kingdom, where organs are almost always accepted by patients towards the very top of the priority list and where mismatch, deterioration, and the associated strategic inter-temporal considerations are thus less relevant, as shown below, allows us to ignore these factors in our analysis and focus on a new aspect of decision-making, which is the ability of centers to correctly assess the quality of the organs they are offered.

Table 1 provides direct evidence that the patient-organ mismatch and organ deterioration, which are both necessarily increasing with a center's position in the queue for a given organ, are less relevant in the United Kingdom. This table reports the relationship between the most stringent (conventional) measure of transplant success – whether the organ survives at least three years – and the position in the queue of the transplanting center, for all livers and kidneys that were transplanted between 2006 and 2015. If mismatch and deterioration are relevant, then organs transplanted at higher positions will have worse outcomes. Because lower quality organs will travel further down the line on average, we include a recently constructed risk index of organ quality (described in greater detail in Section 5.1) in the estimating equation. As observed in Table 1, the probability of transplant success is (not surprisingly) declining significantly in the risk index. Conditional on the risk index, the transplanting center's position in the queue has a negligible effect on transplant success.[4] This indicates that neither mismatch nor deterioration (at higher positions) are relevant in this setting. One alternative explanation for this result is that centers at higher positions account for the mismatch and deterioration in their decision-making; i.e. strategic dynamic considerations are relevant. As we will see in Section

[4]The estimates reported in Table 1 are based on the linear probability model because the marginal effects are easy to interpret. Average marginal effects with the probit model are almost identical to the estimates reported in the table. The estimated marginal effects imply that a two standard deviation increase in center position would reduce the survival probability for both livers and kidneys by 0.02.

Table 1: Determinants of Transplant Success

Dependent variable:	organ survives for at least three years	
Organ:	liver	kidney
	(1)	(2)
Organ risk index	-0.0445***	-0.161***
	(0.010)	(0.020)
Position in queue	-0.00698*	-0.00459
	(0.004)	(0.003)
Center ability	-0.0297	-0.0405
	(0.082)	(0.036)
Mean of dependent variable	0.730	0.787
N	6143	8744

Note: heteroscedasticity-robust standard errors in parentheses
Center ability is the measure used in the herding tests
Position in queue ranges from 1 to 8
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.3, however, there is no evidence that this is the case.

While centers may differ in their ability to distinguish between good and bad organs, we assume that this particular dimension of ability is independent of their competence in implementing transplant procedures. The estimating equation in Table 1 includes the measure of center ability that we will use in the herding tests, which is based on the center's rejection rate when in first position. Although this measure will have a strong effect on decisions, we see (conditional on the risk index) that it has no impact on transplant success. This result implies that higher ability centers, as we define them, do not have greater competence in implementing transplants *and* that their patients do not differ with respect to fixed characteristics such as age and health condition that independently determine transplant outcomes.[5] The results in Table 1, taken together, will allow us to simplify both the model and the empirical analysis that follows.

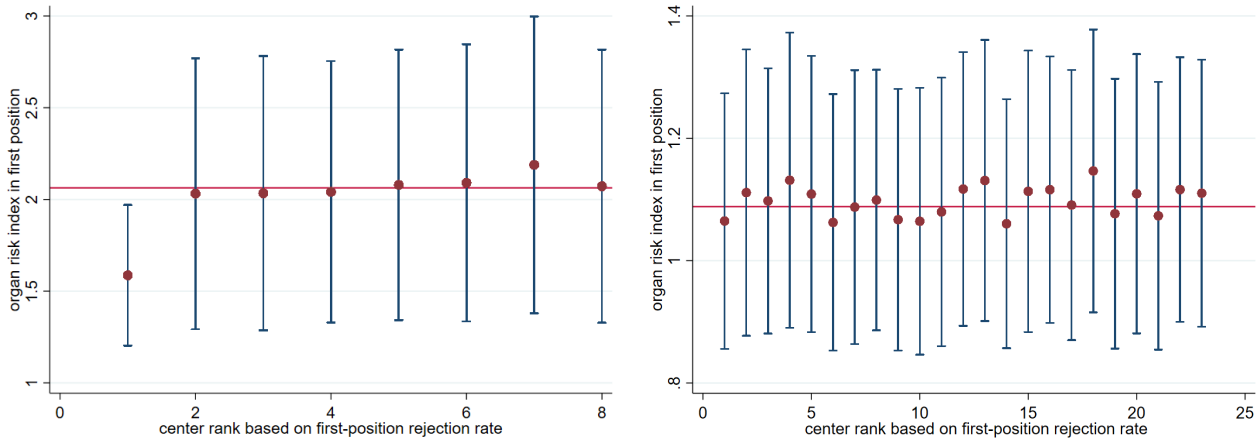
As noted, organs are offered to patients on the basis of their priority, or the proximity of their center to the donor hospital. In particular, center characteristics play no role in the ordering. This feature of the allocation scheme will be especially useful in the analysis that follows and we close this section by verifying two of its implications: (i) The average quality of organs received in first position

[5]Transplant centers service precisely defined regions. Given the length of time that patients must wait for a transplant and the fact that decisions must be made and the transplant itself must be undertaken within a matter of hours, selective sorting by patients into particular centers is unlikely to be a consideration in practice. However, variation in demographic characteristics across regions could, in principle, generate variation in patient characteristics across centers. The results in Table 1 indicate that this is not the case.

does not vary across centers. (ii) The average position at which centers receive organs does not vary across centers.

The X axis in Figure 2 ranks centers by their rejection rate when in first position and the Y axis reports the quality of organs received by centers in that position, where quality is measured by the independently constructed risk index. For each center, the circle marks the mean risk index and the vertical line demarcates one standard deviation above and below the mean. The horizontal line marks the sample mean of the risk index, across all organs in the data. There are eight liver transplant centers and 24 kidney transplant centers and we see that the center-specific means are very close to this population statistic, with the exception of one outlying center for livers.[6] Variation in the rejection rate in first position, which sorts centers on the X axis, is evidently not being driven by variation in the (average) quality of organs they receive, and we will return to this point below.

Figure 2: Risk Index in First Position Across Centers



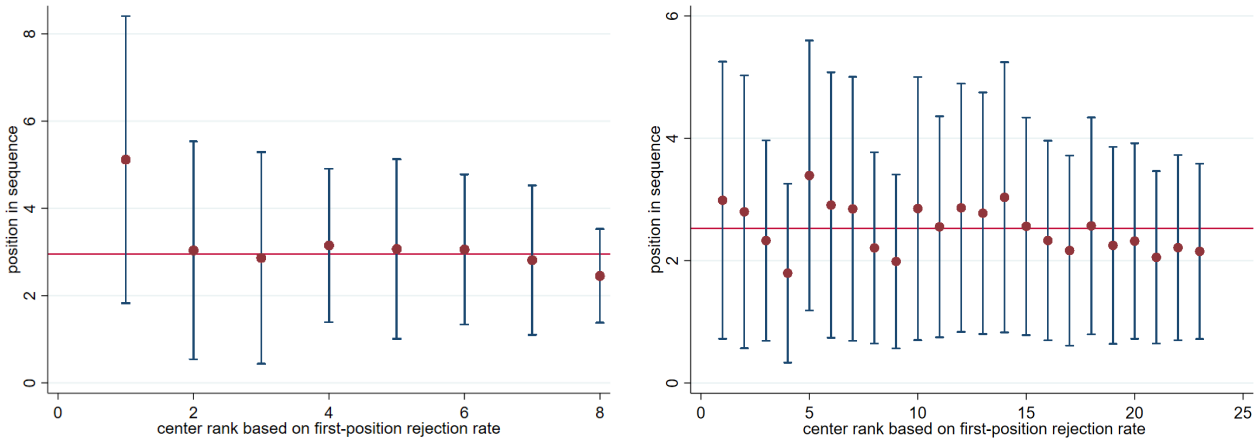
(a) liver (b) kidney

Note: Circles mark the mean risk index for organs received in first position by center and vertical lines demarcate one standard deviation above and below the center-specific mean. The horizontal line marks the sample mean.

Figure 3 replaces organ quality with the center's position in the queue on the Y axis. We see that the center-specific mean positions are, once again, close to the sample mean, which is approximately 3 for livers and 2.5 for kidneys, with the exception of the same outlying liver center. Based on the model that follows, we will associate the rejection rate in first position (the X variable) with center ability and, hence, Figure 3 is consistent with the assumption that centers are ordered independently of their ability. Additional support for this assumption is provided in Section 4.1.

[6]The same outlying center is dropped from Figure 2(b) and from Figure 3(b) that follows because it only receives kidneys in first position. It will also be dropped later in Figures 8 and 9, for kidneys, because the rejection rate in first position lies well below the reported range. We retain this center in the statistical analysis for completeness, but verified that the results of the tests that follow would be unchanged if it were dropped.

Figure 3: Position in the Queue Across Centers



(a) liver (b) kidney

Note: Circles mark the mean position for each center across all the organs it received and vertical lines demarcate one standard deviation above and below the center-specific mean. The horizontal line marks the sample mean.

3 A Model of Organ Transplantation

3.1 Organs, Centers and Signals

Organs can be either of good (G) or bad (B) quality. The outcome of an organ transplant is denoted by H if the organ is good, and by L if it is bad, with $H > 0 > L$. We normalize the outcome of not transplanting an organ to 0. Although centers do not know the quality of a particular organ with certainty, the fraction of organs that are G organs, π, is common knowledge. We define the *cut-off* belief $\tilde{\pi}$ as the belief at which every center is indifferent between accepting or rejecting an organ; i.e. $\tilde{\pi}H + (1-\tilde{\pi})L = 0$, or $\tilde{\pi} = \frac{-L}{H-L}$. $\tilde{\pi}$ could potentially vary across centers because some are more competent at performing transplants; i.e. have higher H/L, due to heterogeneity in patient demographics, or because some centers are more picky than others. As discussed in Section 4.4, allowing for center heterogeneity on this dimension does not undermine our tests of herding.

Centers independently assess organ quality before making a decision. This assessment is based on each center's past experience and the organ-specific information it receives from the NHSBT administrator. More able centers are better equipped to acquire salient information from the administrator and to utilize that information. We characterize each center's independent assessment of an organ by a private information signal $s \in \{g, b\}$, where a g signal indicates that the organ is good, while b indicates that it is bad. We denote a center by j and its ability by $q_j \in \left[\underline{q}, \overline{q}\right] \subset \mathbb{R}$. A center's ability determines the probability γ_j with which it correctly identifies a G organ and the probability β_j with which it correctly identifies a B organ. In particular, for any center j, $P(g \mid G) = \gamma_j = \gamma(q_j)$ and $P(b \mid B) = \beta_j = \beta(q_j)$, where γ, β are strictly increasing functions. In the discussion that follows, we will sometimes refer to q_j as composite ability to distinguish it from center-specific and organ-quality specific abilities, β_j and γ_j.

Centers are not systematically misinformed; i.e. each center is more likely to receive a b (g) signal with a bad (good) organ. This requires:

Assumption 1: *For all* j, $\beta_j \geq 1 - \gamma_j$.

This assumption will certainly be satisfied if the lowest ability center, with ability $\underline{q}$, is completely uninformed; i.e. is equally likely to receive a b (g) signal with a G organ or with a B organ. This implies $\beta(\underline{q}) = 1 - \gamma(\underline{q})$.

Centers in first position update their prior belief that an organ is good, π, upon receiving their signals. Formally, center j's posterior belief that an organ is good, upon receiving signals g and b in first position is denoted by

$$\begin{aligned} \pi_j(g) &= \frac{\pi\gamma_j}{\pi\gamma_j + (1-\pi)(1-\beta_j)} \\ \pi_j(b) &= \frac{\pi(1-\gamma_j)}{\pi(1-\gamma_j) + (1-\pi)\beta_j}, \end{aligned}$$

respectively. Given Assumption 1, the belief shifts up (down) upon receipt of a $g(b)$ signal:

$$\pi_j(g) \geq \pi \geq \pi_j(b). \tag{1}$$

In addition to Assumption 1, we further assume that centers always follow their own signals in first position (absent any other information) such that each center accepts the organ if it receives a g signal and declines the organ upon receipt of a b signal. This is equivalent to the following:

Assumption 2: *For all* j, $\pi_j(g) \geq \tilde{\pi} > \pi_j(b)$.

We verify each of these assumptions in Section 5.2.

3.2 Transplant Decisions

Organs are offered sequentially to centers on the basis of a predetermined algorithm. The priority list for a given organ is based on recipients' characteristics, their match with the organ and, possibly, the distance between their center and the donor hospital. Center ability, in particular, plays no role in this ordering. This is consistent with the observation from the preceding section that organ quality in first position and average position do not vary across centers. Moreover, as discussed in that section, in the institutional environment that we study, where organs rarely proceed past the first few positions, there is little variation in patient characteristics, the organ-patient mismatch, and organ deterioration. Although we will consider a potential role for these factors when interpreting the empirical results, this allows us to focus on centers' beliefs about organ quality when modeling decision-making.

We next describe the evolution of beliefs, and associated decisions, for the centers in line for a given organ. To simplify notation for the rest of this section and for the tests of herding in Section 4, we identify a center by its position in line, such that the center at position j has ability q_j. Center

1 receives a signal and, given Assumption 2, accepts after a g signal and declines after a b signal. If the organ is accepted, it is transplanted by center 1 and results in payoff H or L, depending on its quality. If it is declined, an administrator from NHSBT decides either to offer the organ to the next center or to set it aside for research. The decision to discard an organ is based on its condition, which depends on its quality and its deterioration over the course of the offering process. Although we do not model the discard decision, we will account for its exogenous (common) effect on center beliefs at each position in the empirical analysis, where relevant. Note that the discard decision has no bearing on our tests of herding because they are based on variation in center beliefs and decisions at the same position.

Centers positioned further along in the sequence learn from the (rejection) decisions of their predecessors. Each center knows the identity of its predecessors and the order in which they made their decisions. If this were not the case, then all the tests reported below would fail to be supported by the data.[7] We use an iterative process to describe centers' equilibrium beliefs and strategies moving down the line. The equilibrium concept that characterizes the learning process is Perfect Bayesian Nash Equilibrium.

If center 2 is offered an organ, it knows that center 1 must have received a b signal, given Assumption 2. Its prior belief before it receives its private signal, which is public information, is denoted by π_2; thus

$$\pi_2 = \pi_1(b) = \frac{\pi(1-\gamma_1)}{\pi(1-\gamma_1)+(1-\pi)\beta_1}. \tag{2}$$

Its posterior belief, upon receiving signals g and b, respectively, is given by

$$\pi_2(g) = \frac{\pi_2\gamma_2}{\pi_2\gamma_2+(1-\pi_2)(1-\beta_2)},$$

and

$$\pi_2(b) = \frac{\pi_2(1-\gamma_2)}{\pi_2(1-\gamma_2)+(1-\pi_2)\beta_2}.$$

Center 2 always rejects the organ if it receives a b signal, because its prior belief, π_2 (which is lower than $\tilde{\pi}$ from Assumption 2), is downgraded even further following a b signal. Center 2 could reject the organ even if it receives a g signal – which implies that it is herding – if this updating does not raise its posterior above $\tilde{\pi}$. To summarize, center 2's optimal decision is to accept the organ if it received a g signal and $\pi_2(g) \geq \tilde{\pi}$, and to decline otherwise.

Next, center 3 knows center 2's decision-making process and its prior belief, π_2, but does not necessarily know center 2's signal. If center 2 herds, its decision provides no information about its signal to center 3, and the latter's public belief π_3 is therefore equal to π_2. If, on the other hand, center 2 uses its signal to make its decision ($\pi_2(g) \geq \tilde{\pi}$), then center 3 infers from center 2's rejection

[7] In other settings with sequential decision-making, it is possible that all predecessors' identities will not be available. A modified analysis, on the lines of Çelen and Kariv (2004) will then be required.

that it must have received a b signal, and therefore has a public belief equal to $\pi_2(b)$. Thus,

$$\pi_3 = \begin{cases} \pi_2, & \text{if } \pi_2(g) < \tilde{\pi} \\ \pi_2(b) & \text{otherwise.} \end{cases}$$

The preceding discussion can be easily generalized. In the same way as center 2, center $n > 3$, given its public belief (π_n), forms its posterior belief (either $\pi_n(g)$ or $\pi_n(b)$), and then chooses optimally either to accept or to decline the organ. Then, as with center 3, center $n+1$'s public belief π_{n+1} equals π_n if center n herds and $\pi_n(b)$ otherwise.

4 Herding Tests

4.1 Center Ability

As discussed, the novelty of our tests of herding is that they are based on variation in center decisions at the same position. This variation arises because centers differ in their ability to distinguish between good and bad organs and because the order of centers varies from one organ to the next. The first step in deriving our tests is thus to use the model to construct a composite measure of center ability.

Our measure of a center's ability is based on its observed rejection rate when in first position $\overline{p}_1$. By Assumption 2, which we verify in Section 5.2, all centers follow their signals in first position. Thus, for any specific organ, the probability that center 1 with ability q_1, rejects that organ, $p_1(q_1)$, is equal to the probability that it receives a b signal:

$$p_1(q_1) = \pi(1-\gamma_1) + (1-\pi)\beta_1. \tag{3}$$

Under the assumption that all centers receive organs of the same average quality in first position, as verified in Figure 2, this probability depends only on the probability that the organ is good π and the ability of the center q_1. Moreover, it is the same for all organs that the center receives in first position. This implies that $p_1(q_1) = \overline{p}_1$, and we will thus use these terms interchangeably.

By equation (3), we also have that

$$\frac{dp_1(q_1)}{dq_1} = -\pi\gamma'(q_1) + (1-\pi)\beta'(q_1). \tag{4}$$

It is evident from equation (4) that $p_1(q_1)$ could be increasing or decreasing in q_1 because more able centers are better at detecting both good and bad organs; i.e. $\gamma'(q_1)$ and $\beta'(q_1)$ are both positive. In an inferior organ pool, with many bad organs, the $\beta'(q_1)$ term dominates and $p_1(q_1)$ is increasing in q_1. In a superior organ pool, with many good organs, the $\gamma'(q_1)$ term dominates and $p_1(q_1)$ is decreasing in q_1. We allow for both possibilities, with the restriction that the probability of rejection in first position, for a given organ pool, is either monotonically increasing or decreasing in ability for

all centers; i.e. either $\frac{dp_1(q_1)}{dq_1} > 0$ for any $q_1 \in [\underline{q}, \overline{q}]$ or $\frac{dp_1(q_1)}{dq_1} < 0$ for any $q_1 \in [\underline{q}, \overline{q}]$. We provide empirical support for this monotonicity assumption in Section 5.2.

To determine whether the probability of rejection in first position is increasing or decreasing in center ability, we examine the decisions of centers in second position. To begin with, assume that center 2 follows its own signal. This would be the case if it ignores its predecessors' decisions or if it does not herd; that is, its posterior belief upon receiving a g signal exceeds $\tilde{\pi}$. In this case, the probability that center 2 rejects an organ is the probability that it receives a b signal, conditional on center 1 also having received a b signal. Because signals are conditionally independent and centers follow their own signals in first position (Assumption 2), the probability that center 2 rejects an organ conditional on center 1 rejecting is thus given by

$$p_2(q_1, q_2) \equiv \frac{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{5}$$

We can then compute the manner in which center 2's rejection probability, which we also refer to as p_2 in the discussion that follows, varies with center 1's ability:

$$\frac{\partial p_2(q_1, q_2)}{\partial q_1} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2 - (1-\gamma_2))}{(\pi(1-\gamma_1) + (1-\pi)\beta_1)^2} \geq 0. \tag{6}$$

The inequality in expression (6) follows from Assumption 2. It implies that if center 2 follows its own signal, then it is more likely to reject when its predecessor has higher ability. Moreover, an increase in center 1's ability makes it more likely that center 2 herds, in which case it rejects for sure (regardless of its signal). This is because a higher-ability predecessor's rejection has a bigger impact on center 2's prior belief, thereby increasing the likelihood that its posterior belief will remain below $\tilde{\pi}$ even when it receives a g signal. Thus, center 2 is more likely to reject an organ when center 1 has high ability, regardless of whether centers learn from their predecessors or not.

Based on the preceding discussion, if we observe that a center in second position is more (less) likely to reject an organ when it follows a center with a higher rejection rate in first position, $\overline{p}_1$, then $\overline{p}_1$ or, equivalently, $p_1(q_1)$, must be positively (negatively) associated with center ability. If center 2 ability, q_2, and center 1 ability, q_1, are correlated, and q_2 is omitted from the estimating equation, then $\overline{p}_1$ will proxy for both q_1 and q_2. Based on the evidence reported in Figure 3 we do not expect this to be the case, but we nevertheless include center 2 fixed effects in all specifications in the table that follows.

The relationship between the decision in second position for each organ that reaches that position and center 1's rejection rate when in first position, $\overline{p}_1$, is reported in Table 2, Columns 1 and 3, for livers and kidneys respectively.[8] We also report results with an augmented specification that includes

[8]We use the linear probability model to estimate our measures of ability and for the tests of herding because it follows directly from the specified estimating equations and because the marginal effects are easy to interpret. However, this implies that the error term will be heteroscedastic and, hence, robust standard errors are reported in this table and the tables that follow.

the mean organ risk index for center 1 in first position, $\overline{R}_1$, in Columns 2 and 4. The coefficient on $\overline{p}_1$ is positive and significant for livers, and negative and significant for kidneys, with both specifications. Our interpretation of these results is that livers (kidneys) are drawn from inferior (superior) organ pools. Higher ability centers thus reject more often in first position with livers and less often with kidneys.

Table 2: Measuring Center Ability

Dependent variable:	decision in second position			
Organ:	liver		kidney	
	(1)	(2)	(3)	(4)
$\bar{p}_1$	0.513***	0.481***	-0.411***	-0.395***
	(0.071)	(0.070)	(0.037)	(0.037)
$\bar{R}_1$	No	Yes	No	Yes
Center 2 fixed effects	Yes	Yes	Yes	Yes
Mean of dependent variable	0.816	0.816	0.710	0.710
N	6383	6380	9257	9257

Note: heteroscedasticity-robust standard errors in parentheses
Decision in second position: reject = 1, accept = 0
$\bar{p}_1$ measures center 1's rejection rate in first position
$\bar{R}_1$ measures mean organ risk index for center 1 in first position
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The implicit assumption underlying this interpretation is that average organ quality in first position does not vary across centers. If the rejection rate in first position, $\overline{p}_1$, varies (in part) because centers receive organs of different quality on average, then $\overline{p}_1$ will proxy for both q_1 and organ quality (which directly determines decisions at position 2). Based on the evidence presented in Figures 2, we do not expect this source of bias to be present, which implies that the estimated $\overline{p}_1$ coefficient should be stable when $\overline{R}_1$ is included in the estimating equation, and this is indeed the case. Variation in $\overline{p}_1$ can be attributed to differences in center ability and we consequently measure center ability by $\overline{p}_1$ for livers and by $1 - \overline{p}_1$ for kidneys in the tests of herding that follow.

Although we specify that $\tilde{\pi}$ is constant in the model, it is possible that centers vary in their (unobserved) pickiness. This will generate variation in $\overline{p}_1$ across centers that is independent of ability, as we have defined it. We provide additional support for the assumption that $\overline{p}_1$ is a valid proxy for center ability in the following ways: First, we show in Section 4.4 that variation in $\tilde{\pi}$, on its own, cannot generate the observed patterns in Table 2. Second, we show in Section 5.2 that this composite ability measure is positively associated with independently constructed organ-quality specific measures of center ability.

Notice that we do not cluster standard errors in Table 2. The dependent variable in Table 2 is the decision by center 2 to reject or accept a given organ and the residual in the estimating equation incorporates the organ-specific signal that center 2 receives. Center 1 must have rejected the organ and, given the assumption (verified below) that all centers follow their signals in first position, must have received a b signal. However, there will be variation in the type of organs that it passes down on account of the mistakes that it makes. While it will often correctly receive a b signal for a B organ, it will also on occasion receive a b signal for a G organ. The organ's quality (B or G) is independent across organs received by center 2 and, as assumed in the model, the signals it receives are independent, conditional on the underlying quality of the organ. It follows that the error (residual) term in the estimating equation is independent across organs for a given center 1-center 2 pair. By the same argument, the error term will be independent across organs with different center-pairs in first and second position. The estimated standard errors in Table 2 should not be clustered and this is also true, for the same reasons, for the tests of herding that follow.[9]

4.2 A Test of Herding (based on decisions in second position)

The discussion in Section 4.1 indicates that center 2 is more likely to reject an organ when it follows a higher-ability center; i.e. $dp_2/dq_1 < 0$, with and without herding. To test for herding in second position we thus need to look more closely at the $p_2{-}q_1$ relationship, and we do this by examining how this relationship varies with q_2. In particular, we estimate the following equation: $p_2 = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_1 \cdot q_2$, and focus attention on the interaction term.

To begin with, assume that center 2 does not herd, such that its rejection probability for a given organ is described by equation (5). The cross-partial with respect to q_1 and q_2, which is essentially the coefficient on the interaction term, is then

$$\frac{\partial^2 p_2(q_1, q_2)}{\partial q_1 \partial q_2} = \frac{\pi(1-\pi)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1))(\beta_2' + \gamma_2'))}{(\pi(1-\gamma_1) + (1-\pi)\beta_1)^2} > 0, \tag{7}$$

That is, the effect of an increase in center 1's ability on center 2's rejection probability is larger when center 2 has higher ability. This result is obtained because (i) an increase in q_1 reduces the quality of the organ pool passed on to center 2, i.e. the prior belief π_2 goes down and (ii) center 2's decision, p_2, is more sensitive to π_2 when it has higher ability. In the extreme case, if center 2 is completely uninformed; i.e. $\beta_2 = 1 - \gamma_2$, then its decision will be unaffected by the change in the quality of the organ pool.

It follows from equation (7) that the cross partial is positive in the absence of herding. This result, however, does not necessarily hold when center 2 herds. In particular, there are now two effects: the *quality selection effect* and the *herding effect*. The first effect, which we described above, implies that

[9] Abadie et al. (2023) note that if the researcher assesses that the assignment mechanism is not clustered, as we do based on the model and the institutional setting, then the standard errors should not be clustered. This is true irrespective of whether such an adjustment would change the standard errors.

a low-ability center reacts less to an increase in its predecessor's ability than does a high-ability center, because the former is less sensitive to the quality of its organ pool. The second effect works in the opposite direction; as the rejection by a high-ability center 1 represents worse news about underlying organ quality than does the rejection of a low-ability center, it is more likely that a weak center 2 herds and also rejects. Which effect dominates depends on the underlying organ pool and on the ability of the relevant centers. This is best seen in the following two figures. In drawing these figures, we have assumed that the lowest-ability center is completely uninformed and that the highest-ability center is perfectly informed, although the results do not rely on those assumptions; i.e. we assume $\beta(\underline{q}) = 1 - \gamma(\underline{q})$ and $\beta(\overline{q}) = \gamma(\overline{q}) = 1$. The figures show the rejection probabilities of two centers at position 2 as a function of center 1's ability q_1. The abilities of the two centers are q_2' and q_2''; and we assume that $q_2' < q_2''$.

The curves labelled $p_2(\cdot, q_2')$ and $p_2(\cdot, q_2'')$ in the figures are the centers' respective rejection probabilities *without herding*, given by the expression in equation (5). If center 2 follows center 1 with $q_1 = \underline{q}$ it faces an organ pool with quality π and, consequently, draws signals as if it were in first position. This implies that $p_2(\underline{q}, q_2) = p_1(q_2)$, which pins down the intercept of each curve. Note that Figure 4 assumes that $p_1(q_2') < p_1(q_2'')$, which applies to livers as documented above, while Figure 5 assumes that the inequality is reversed, which is relevant for kidneys. This represents the only difference between the two figures. If center 2 follows center 1 with $q_1 = \bar{q}$, the organ is a B organ for certain, so center 2 draws signals from a B organ. This explains why the curves reach heights of $p_2(\overline{q}, q_2') = \beta(q_2')$ and $p_2(\overline{q}, q_2'') = \beta(q_2'')$, where $\beta(q_2') < \beta(q_2'')$.

Figure 4: $\frac{dp_1(q)}{dq} > 0$

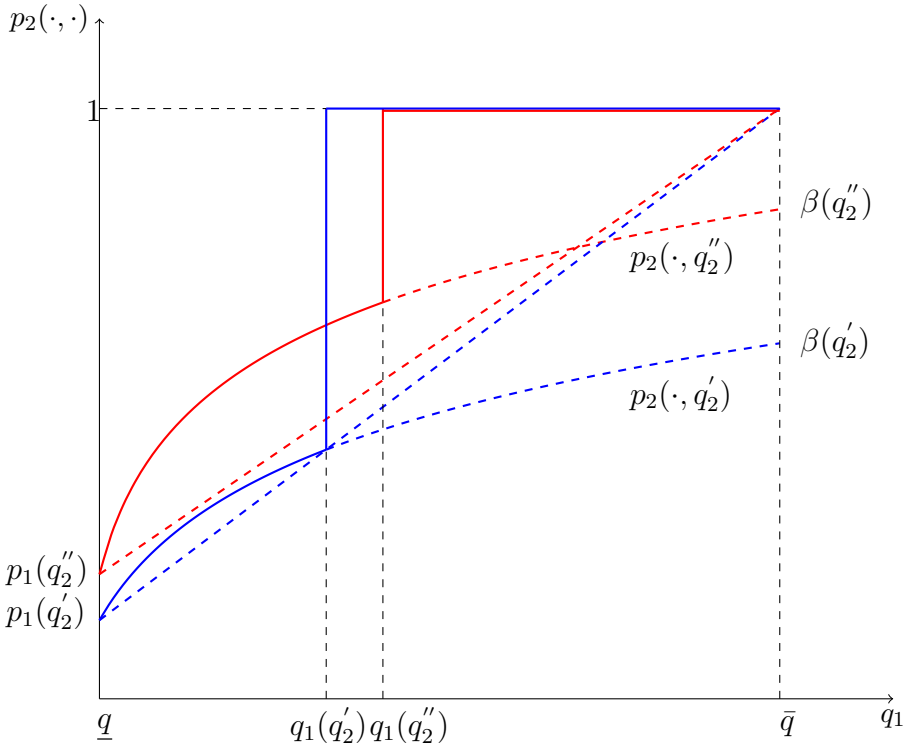
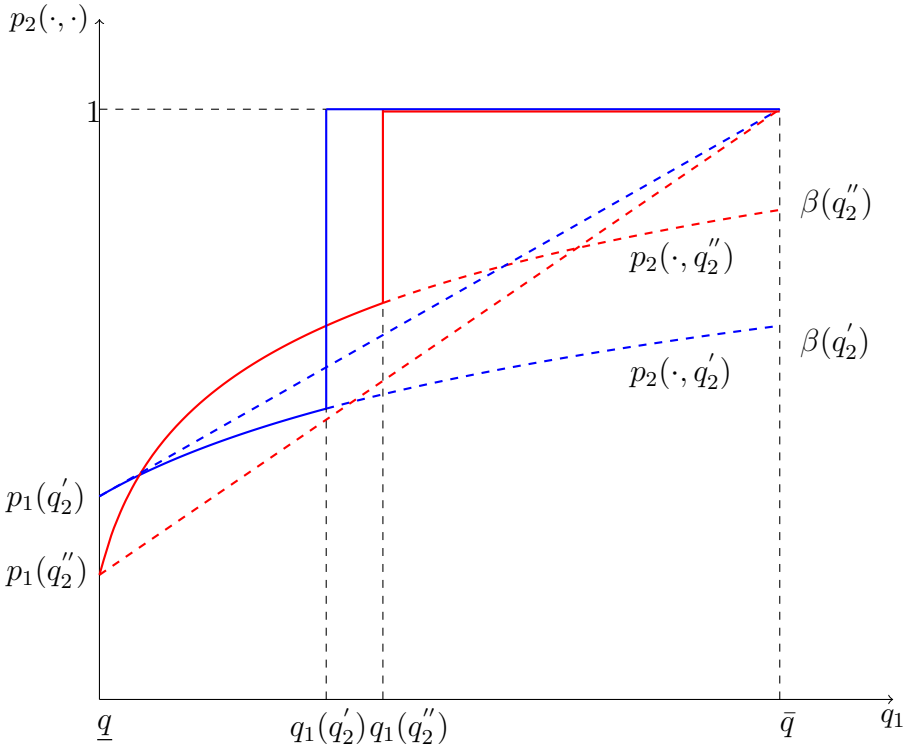


Figure 5: $\frac{dp_1(q)}{dq} < 0$



With herding the rejection probability of any center 2 with ability q_2 jumps to 1 at some threshold $q_1(q_2)$; thus, for $q_1 < q_1(q_2)$, center 2 uses its signal and rejects according to the expression in equation (5) while, for $q_1 \geq q_1(q_2)$, center 2 herds and always rejects. Note that $q_1(q_2') < q_1(q_2'')$ as a lower-ability center positioned at 2 starts herding sooner than does a high-ability one.

We now use the figures to derive the effect of the interaction term $q_1 \cdot q_2$, on center 2's rejection probability with herding. Expression (7) allows us to derive this effect for each pair (q_1, q_2) in the absence of herding. With herding, we cannot use the same method because center 2's rejection probability contains a jump when center 2 starts to herd and, hence, the derivative is not well-defined at each point. Instead, for each q_2, we compute the"average effect" of an increase in q_1; this is the slope of the line starting at $(\underline{q}, p_1(q_2))$ and going to $(\bar{q}, \beta(q_2))$ if there is no herding and to $(\bar{q}, 1)$, otherwise.[10] We then examine how this slope varies with q_2 (q_2' versus q_2'').

Consistent with the cross-partial expression in (7), the average slope with respect to q_1 is increasing in q_2 in the absence of herding in both figures:

$$\frac{\beta(q_2') - p_1(q_2')}{\bar{q} - \underline{q}} < \frac{\beta(q_2'') - p_1(q_2'')}{\bar{q} - \underline{q}}$$

When we incorporate the effect of herding, however, we see that the slope with respect to q_1 in Figure

[10]The implicit assumption when computing the average effect is that the distribution of q_1 is uniform on $[\underline{q}, \bar{q}]$.

4 is decreasing in q_2:

$$\frac{1-p_1(q_2')}{\bar{q}-\underline{q}} > \frac{1-p_1(q_2'')}{\bar{q}-\underline{q}}.$$

In contrast, the slope with respect to q_1 in Figure 5 is increasing in q_2. Thus, when center ability is *increasing* in the probability of rejection in first position, $\left(\frac{dp_1(q)}{dq} > 0\right)$, as observed in Figure 4 and for livers, we predict that the interaction effect is reversed with herding: lower-ability centers at position 2 react more to an increase in center 1's ability because the herding effect dominates. When center ability is *decreasing* in the probability of rejection in first position, $\left(\frac{dp_1(q)}{dq} < 0\right)$, as observed in Figure 5 and for kidneys, we predict that the interaction effect goes in the same direction as it does in the absence of herding; the quality selection effect then dominates, such that higher-ability centers at position 2 react more to an increase in center 1's ability. We summarise the preceding discussion as follows:

Test 1 *Without herding, the cross-partial effect of an increase in both center 1's and center 2's ability on center 2's rejection probability, for a given organ, is strictly positive. With herding, the average cross-partial effect is strictly positive if center ability is decreasing in the rejection rate in first position and negative if center ability is increasing in the rejection rate in first position.*

Table 3 reports the estimated relationship between the decision in second position for each organ that reaches that position and center ability in first and second position, together with the interaction term. Center ability is measured by the rejection rate in first position for livers and by one minus that statistic for kidneys. The same measures are used for the test of herding at third position that follows.[11]

To interpret the estimated coefficients, it is convenient to normalize so that $\underline{q} = 0$. α_1, the coefficient on q_1, then applies to the case where q_2 equals zero. Assume that a center with ability 0 is completely uninformed; i.e. $\beta(0) = 1 - \gamma(0)$. This implies that without herding, a center with $q_2 = 0$ makes decisions that are independent of the quality of the organs that it receives and, hence, are independent of q_1. However, with herding, the probability that such a center rejects for sure is increasing in q_1. The predicted effect is ambiguous, and we find that α_1 is small and imprecisely estimated for kidneys and much larger and significant at the one percent level for livers. In line with these results, we will see below that the prevalence of herding in second position is substantially higher for livers than for kidneys.

α_2, the coefficient on q_2, applies to the case in which q_1 equals zero. In this case, the first center's decision has no effect on the quality of the organ pool that is passed on and center 2 effectively behaves as if it is in first position. We thus expect higher ability (q_2) centers to reject more often for livers

[11] The implicit assumption in the model is that no center decides more than once for a given organ. Given that centers have multiple patients on the waiting list, it is possible that this requirement will not be satisfied in practice. It turns out that sequences in which the same center makes repeated decisions are rare in the data; 7% of all decisions for kidneys and 3% of all decisions for livers are repeat decisions made by a center with the same organ. Restricting attention to the first three positions, which we use for the herding tests, these statistics decline even further to 5% and 2%, respectively.

and less often for kidneys. As predicted, the coefficient on q_2 is positive and significant for livers and negative and significant for kidneys. Our test of herding, however, is based on the interaction coefficient. As predicted by the model, the interaction coefficient is negative and significant for livers, and positive and significant for kidneys in Columns 1 and 3.[12] This is also true with a more flexible specification in Columns 2 and 4, which includes center 1 and center 2 fixed effects to account for systematic variation in organ quality or the organ-patient mismatch across centers. In contrast with these results, the interaction coefficient would be positive and significant for both livers and kidneys in the absence of herding.

Table 3: First Test of Herding (based on decisions in second position)

Dependent variable:	decision in second position			
Organ:	liver		kidney	
	(1)	(2)	(3)	(4)
Center 1 ability (q_1)	2.135**	–	-0.0269	–
	(0.713)		(0.097)	
Center 2 ability (q_2)	2.588***	–	-1.000***	–
	(0.688)		(0.102)	
($q_1 \times q_2$)	-2.668**	-3.475***	1.009***	1.226***
	(1.103)	(0.998)	(0.246)	(0.255)
Center 1 fixed effects	No	Yes	No	Yes
Center 2 fixed effects	No	Yes	No	Yes
Mean of dependent variable	0.816	0.816	0.710	0.710
N	6383	6383	9257	9257

Note: heteroscedasticity-robust standard errors in parentheses.
Decision in second position: reject = 1, accept = 0
The constant term cannot be interpreted and is thus not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.05$

4.3 A Test of Herding (based on decisions in third position)

Our second test of herding is based on decisions at position 3; in particular, on the relationship between these decisions and center abilities at position 1 (q_1) and position 2 (q_2), as expressed in the following equation: $p_3 = \lambda_0 + \lambda_1 q_1 + \lambda_2 q_2$. In deriving this test we assume that center 3 does not herd - centers

[12]Note that the results in Table 3 are consistent with the key finding from Table 2, where the specification does not include an interaction term, which is that the probability of rejection in second position, p_2, is increasing (decreasing) in center 1's rejection rate in first position, $\overline{p}_1$, with livers (kidneys). Based on the point estimates in Table 3, p_2 is increasing (decreasing) in $\overline{p}_1$ with livers (kidneys) for all values of q_2 in the data.

that herd at third position always reject, and their decision is thus unaffected by marginal changes in q_1 and q_2.

To develop our second test of herding we investigate the effect of a marginal increase in q_1 and q_2 on center 3's rejection probability p_3. We first consider the case in which center 2 does not herd; then p_3 is the probability that center 3 receives a b signal, conditional on both center 1 and center 2 also having received b signals. In this case p_3 is given by

$$p_3(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_2)(1-\gamma_3) + (1-\pi)\beta_1\beta_2\beta_3}{\pi(1-\gamma_1)(1-\gamma_2) + (1-\pi)\beta_1\beta_2}. \tag{8}$$

It is easy to see that an increase in either center 1's or center 2's ability decreases the quality of the organ pool passed on to center 3, which increases the latter's rejection probability. Formally,

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_1} = \Theta\beta_2(1-\gamma_2)(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)), \tag{9}$$

$$\frac{\partial p_3(q_1, q_2, q_3)}{\partial q_2} = \Theta\beta_1(1-\gamma_1)(\gamma_2'\beta_2 + \beta_2'(1-\gamma_2)), \tag{10}$$

where, $\Theta = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)(1-\gamma_2)+(1-\pi)\beta_1\beta_2]^2}$. Both expressions clearly are strictly positive, but their exact magnitudes depend on the abilities of centers 1 and 2. When $q_1 \neq q_2$, $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_1}$ could be larger or smaller than $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_2}$; for $q_1 = q_2$, however, $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_1}$ is equal to $\frac{\partial p_3(q_1,q_2,q_3)}{\partial q_2}$.

Now consider a situation in which center 2 herds. Because it rejects, regardless of the signal it receives, its decision has no effect on the quality of the organ pool passed on to center 3. In contrast, center 1 always uses its signal. Center 3's rejection probability, p_3, is thus the probability that center 3 receives a b signal, conditional only on center 1 having received a b signal:

$$p_3^h(q_1, q_2, q_3) = \frac{\pi(1-\gamma_1)(1-\gamma_3) + (1-\pi)\beta_1\beta_3}{\pi(1-\gamma_1) + (1-\pi)\beta_1}. \tag{11}$$

In this case, the effect of an increase in the ability of center 1, though different to the case without herding (see equation (9)), is still positive, while the effect of an increase in the ability of center 2 is zero:

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_1} = \Pi(\gamma_1'\beta_1 + \beta_1'(1-\gamma_1)) \tag{12}$$

$$\frac{\partial p_3^h(q_1, q_2, q_3)}{\partial q_2} = 0, \tag{13}$$

where $\Pi = \frac{\pi(1-\pi)(\beta_3-(1-\gamma_3))}{[\pi(1-\gamma_1)+(1-\pi)\beta_1]^2}$. Summarizing the preceding discussion:

Test 2 *If centers* 1 *and* 2 *have the same ability then, without herding, the effect of an increase in center* 1*'s ability on center 3's rejection probability equals the effect of an increase in center 2's ability. With herding, the effect of an increase in center* 1*'s ability is larger than the effect of an increase in center* 2*'s ability.*

Table 4 reports the estimated relationship between the decision in third position for each organ that reached that position, and center abilities q_1 and q_2. As predicted by the model when herding is present, the coefficient on q_1 is substantially larger than the coefficient on q_2; it is twice as large for livers and 50% larger for kidneys.[13] The coefficients on q_1 and q_2 (λ_1 and λ_2 respectively) are imprecisely estimated for livers, and we cannot reject the hypothesis that $\lambda_1 \leq \lambda_2$. The corresponding coefficients for kidneys are, however, statistically significant; we can reject the hypothesis that $\lambda_1 \leq \lambda_2$ at the 5 per cent level.

Table 4: Second Test of Herding (based on decisions in third position)

Dependent variable:	decision in third position	
Organ:	liver	kidney
	(1)	(2)
Center 1 ability (q_1)	0.104	0.352***
	(0.066)	(0.045)
Center 2 ability (q_2)	0.0529	0.220***
	(0.064)	(0.046)
Constant	0.820***	0.541***
	(0.067)	(0.024)
F-statistic ($\lambda_1 \leq \lambda_2$)	0.47	3.43
p-value	[0.247]	[0.032]
$\bar{q_1}$	0.60	0.40
$\bar{q_2}$	0.65	0.40
N	4819	6084

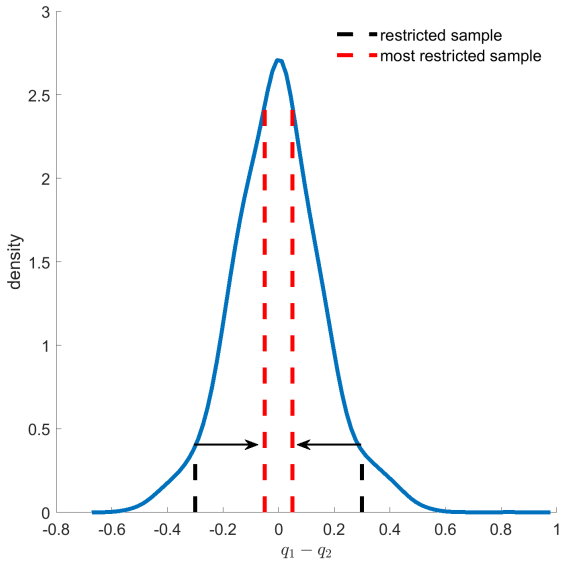
Note: λ_1, λ_2 are the coefficients on q_1, q_2, respectively
$\bar{q_1}$ and $\bar{q_2}$ denote the sample means of q_1 and q_2, respectively.
Heteroscedasticity-robust standard errors in parentheses
Decision in third position: reject = 1, accept = 0
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The data requirements to implement the second test of herding are quite stringent: (i) A substantial fraction of centers should herd in second position. (ii) A substantial fraction of centers should *not* herd in third position (if they did, then variation in q_1, q_2 would have no consequence for their decisions). (iii) There should be substantial variation in decisions – accept versus reject – in third position for the

[13]The identifying assumption when implementing this test is that center 2 does not ignore its (positive) signal for other reasons; for example, because the organ has deteriorated or because the patient-organ mismatch has increased. As discussed in Section 2, we do not expect deterioration or mismatch to be relevant in this setting and this is particularly true when comparing behavior in first and second position.

test to have statistical power. We will see below that conditions (i) and (ii) are satisfied for both livers and kidneys. The important difference between the two organ types is that by the third position, over 90% of decisions for livers are rejections. This lack of variation might explain why the coefficients on q_1 and q_2 are imprecisely estimated in Column 1 of Table 4. While livers are most useful for identifying herding with the first test, we thus focus on kidneys for the second test.

Figure 6: Ability Differential ($q_1 - q_2$) Distribution



Note: sample includes all kidneys that reached third position.

Although Test 1 places no restrictions on center abilities, Test 2 is derived for the case where centers at position 1 and 2 have the same ability; i.e. $q_1 = q_2$. Figure 6 describes the distribution of the ability differential, $q_1 - q_2$, for all kidneys that reached at least third position (and are thus used for the second test of herding). Although the distribution is centered at zero, consistent with the observation in Table 4, Column 2 that the average ability in first position, $\overline{q}_1$, is equal to the average ability in second position, $\overline{q}_2$, there is substantial variation in the ability differential statistic. Table 5 takes account of this variation in $q_1 - q_2$ by implementing the second test of herding with an increasingly restricted sample of kidneys; i.e. by gradually narrowing the ability differential range. Note that restricting the sample in this way does not bias our estimates because center abilities are exogenously assigned. We see that the key result from Table 4, which is that the coefficient on q_1 is significantly larger than the coefficient on q_2 for kidneys, is retained as we reduce the sample. Indeed, this result is even obtained with the most stringent ability-differential restriction in Column 4, by which point the sample is just one-quarter of the full sample of kidneys.

Table 5: Second Test of Herding (restricted samples)

Dependent variable:	decision in third position			
$(q_1 - q_2)$ range:	[-0.30,0.30] (1)	[-0.20,0.20] (2)	[-0.10,0.10] (3)	[-0.05,0.05] (4)
Center 1 ability (q_1)	0.396***	0.427***	0.638***	1.020**
	(0.053)	(0.064)	(0.136)	(0.468)
Center 2 ability (q_2)	0.153**	0.132**	0.00181	-0.437
	(0.055)	(0.065)	(0.140)	(0.470)
Constant	0.554***	0.550***	0.513***	0.515***
	(0.025)	(0.027)	(0.039)	(0.045)
F-statistic ($\lambda_1 \leq \lambda_2$)	7.27	6.81	5.94	2.44
p-value	[0.004]	[0.004]	[0.007]	[0.059]
N	5603	5071	3069	1665

Note: heteroscedasticity-robust standard errors in parentheses

Alternative samples restricted to kidneys within a pre-specified ability differential $(q_1 - q_2)$ range

Decision in third position: reject = 1, accept = 0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Alternative Sources of Heterogeneity

Our analysis incorporates heterogeneity in organ quality and center ability. In this section we examine the possibility that other sources of heterogeneity could bias our estimates of center ability or confound the tests of herding.

In our model, centers make decisions based on their assessment of organ quality alone. In theory, these decisions will also depend on the patient-organ match and the deterioration of the organ (although we have noted that these factors may be less important in practice). These organ-center specific factors are idiosyncratic, and thus the rejection rate across all organs received by a center in first position, $\overline{p}_1$, will continue to serve as a valid proxy for its ability. By the same argument, organ-center specific factors will not confound our tests of herding. These tests are based on decisions at a single – second or third – position, with center 1 and center 2 ability included in the estimating equation. An idiosyncratic organ-center specific factor will appear in the residual of the estimating equation with each test, but it will be orthogonal to q_1 and q_2.

Given that center ability is the source of variation in the herding tests, a more relevant concern is that other dimensions of center heterogeneity will contaminate our measure of ability, based on $\overline{p}_1$, or independently generate the associations that we attribute to herding. We have seen in Figure 2 that average organ quality in first position, measured by the risk index, does not vary across centers.

However, other sources of heterogeneity must also be considered, including (i) that centers differ in their ability (along other dimensions) to perform transplants, or (ii) that centers have the same ability, but some are more conservative than others, or (iii) that patient demographics vary across centers, such that some centers are more picky than others.

The first point to note is that all of the alternative explanations listed above effectively generate variation in $\tilde{\pi}$ across centers. The second point to note is that this variation in $\tilde{\pi}$ directly impacts the center-specific rejection rate in first position. However, if centers do not differ in their ability to detect good and bad organs, then more picky centers with a higher rejection rate (higher $\tilde{\pi}$ and $\overline{p}_1$) will pass on higher quality organs on average to the centers that follow, resulting in a lower probability of rejection in second position. This implies that the $\overline{p}_1$ coefficient in Table 2 will be unambiguously negative, which is at odds with what we observe in the table (for livers). As shown in the Appendix, the preceding argument is robust to alternative signal structures and is obtained regardless of whether or not center 2 learns from its predecessor's decision.[14]

Our first test of herding, based on decisions in second position, is an augmented version of the specification in Table 2 (with an additional interaction term). Thus, if heterogeneity in $\tilde{\pi}$ cannot explain the results in Table 2, then it will not explain the results in Table 3. Moreover, as shown in the Appendix, heterogeneity in $\tilde{\pi}$ cannot explain observed decisions in third position, as reported in Tables 4 and 5; in particular, the differential marginal effect of q_1 and q_2, in the absence of herding. In contrast, our model, which is based on heterogeneity in the ability of centers to detect good and bad organs, can generate the results in Table 2 if the quality of the organ pool varies by organ type. In particular, the model predicts that the $\overline{p}_1$ coefficient will be positive in an inferior (low-π) organ pool, which we infer is the case for livers, and negative in a superior (high-π) organ pool, which we infer is the case for kidneys. Independent evidence presented in Section 5.2 indicates that π is indeed substantially larger for kidneys than for livers. The results in Tables 3-5, based on our interpretation of $\overline{p}_1$ as a measure of center ability, are then consistent with herding. We are unaware of any alternative explanation for these results.

5 Model Estimation and Validation

5.1 Ability Parameter Estimates

In the model, each center j has composite ability, q_j, which is positively associated with its ability to detect bad organs, β_j, and good organs, γ_j. While q_j was sufficient to test for herding, we will need organ-quality and center-specific measures, β_j and γ_j, to verify the assumptions of the model, to examine its goodness of fit, and for the efficiency analysis. To estimate β_j, γ_j, the main challenge is

[14] Suppose that the quality of organs received in first position varies across centers (and is not fully captured by the risk index). If centers do not differ in their ability to detect good and bad organs or in their pickiness ($\tilde{\pi}$), then following the argument above, centers who receive lower quality organs on average in first position will have a higher rejection rate *and* will pass on worse organs on average. This implies that the $\overline{p}_1$ coefficient in Table 2 will be unambiguously positive, which is once again at odds with what we observe (for kidneys).

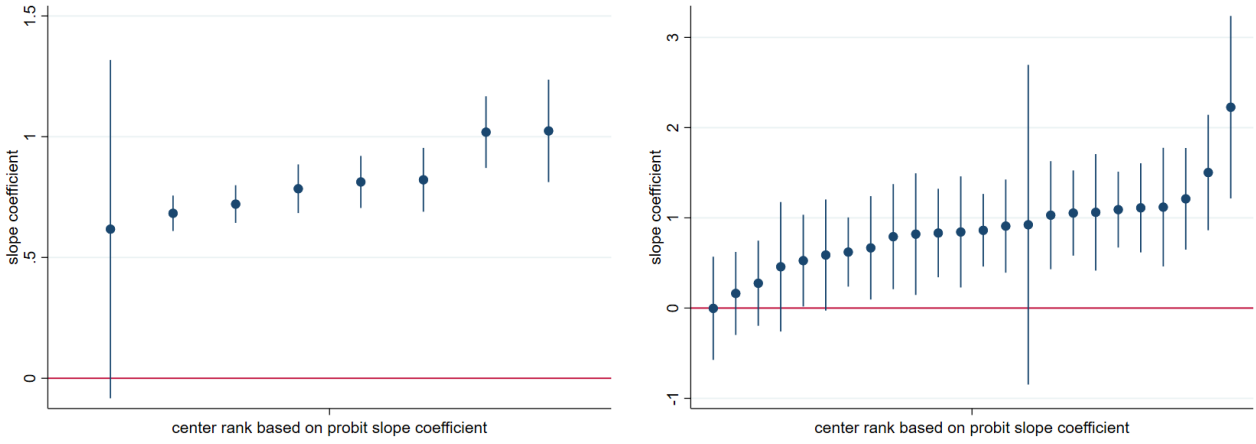
that we do not observe the quality of a given organ in the data; i.e. whether it is good or bad. This is true even *ex post*, especially if the organ is not accepted. A novel feature of our analysis is that we use independently constructed "risk indices" of organ quality, which we interpret through the lens of the model as being negatively associated with the probability that a given organ is a G organ, to estimate β_j, γ_j.

Indices of liver and kidney quality were first constructed in the United States, but have recently been adapted to the U.K. population. The UK KDRI (Kidney Donor Risk Index) is based on U.K. National Transplant Registry data covering over 7000 recipients who received deceased-donor kidneys between January 1, 2000 and December 31, 2007 (Watson et al., 2012). Various recipient and transplant factors were included in a model of transplant success, measured by patient survival, and the UK KDRI consists of those donor and organ characteristics that were found to be significant determinants of success (with optimal, estimated weights on each of those characteristics). More recently, data from all liver transplants from deceased donors between January 1, 2000 and December 31, 2014 have been used to construct the UK DLI (Donor Liver Index). As with the UK KDRI, donor, recipient, and transplant data were used to identify factors associated with graft survival. Those donor and organ characteristics that were found to be significant determinants of transplant success, appropriately weighted, are included in the UK DLI (Collett et al., 2017).

The risk indices are based on outcomes generated by thousands of transplants over many years. The set of organ and donor characteristics included in the indices, and the weights placed on these characteristics, taken together, will accurately predict transplant outcomes or, equivalently, the quality of the organ. In contrast, transplant centers must base their assessment of an organ's quality, g or b, on their past experiences with a limited set of outcomes and the organ-specific information that they receive from NHSBT. The risk indices were originally developed to aid centers in their decision-making, and a proposal to incorporate the UK KDRI into the National Kidney Offering Scheme was presented at the 2018 Blood and Transplantation Congress. At the time of writing, however, neither the UK KDRI nor the UK DLI, the latter of which was developed in 2017, are made available to transplant surgeons when they make their decisions. While centers may not have had explicit knowledge of the risk indices during the period of our analysis (2006-2015), we have assumed that they are not systematically misinformed and, hence, we expect that their assessments and their (rejection) decisions will track with the risk index.

To test the preceding hypothesis, we estimate the relationship between the probability that an organ i is rejected and its risk index R^i, separately by center, restricting the sample to decisions that were made when centers were in first position (and therefore, by assumption, following their signals). We use the probit model for the estimation because this ensures that predicted rejection probabilities lie in the unit interval; these predicted values will be needed to estimate β_j, γ_j, as discussed below. Figure 7 reports probit estimates of the Risk Index (slope) coefficient, with the corresponding 95% confidence interval, by center. The estimated coefficient is positive and statistically significant, almost

Figure 7: Probit Slope Coefficient Estimates Across Centers



(a) liver (b) kidney

Note: estimates based on the relationship between the probability of rejection in first position and the risk index.

without exception, both for livers and kidneys. Notice, however, that there is substantial variation in this coefficient across centers. We would expect rejection decisions by higher ability centers to track more closely with the risk index, and we will build on this intuition to estimate β_j, γ_j below.

Based on our interpretation of the risk index, the probability that organ i is a good organ, π^i, is decreasing in its risk index, R^i. We model this by assuming that $\pi^i = \pi(R^i)$ for some decreasing function $\pi(.)$. We place two additional restrictions on the $\pi(.)$ function: $\pi(\underline{R}) = 1$, $\pi(\overline{R}) = 0$, where $\underline{R}$, $\overline{R}$ define the support of the risk index distribution. This is simply saying that an organ is a G organ with probability one (zero) at the bottom (top) of the risk index distribution. Given that the probability that center j will reject organ i in first position, p^i_{j1}, is equal to the probability that it receives a b signal, the following must hold

$$p^i_{j1} = \pi^i(1-\gamma_j) + (1-\pi^i)\beta_j. \tag{14}$$

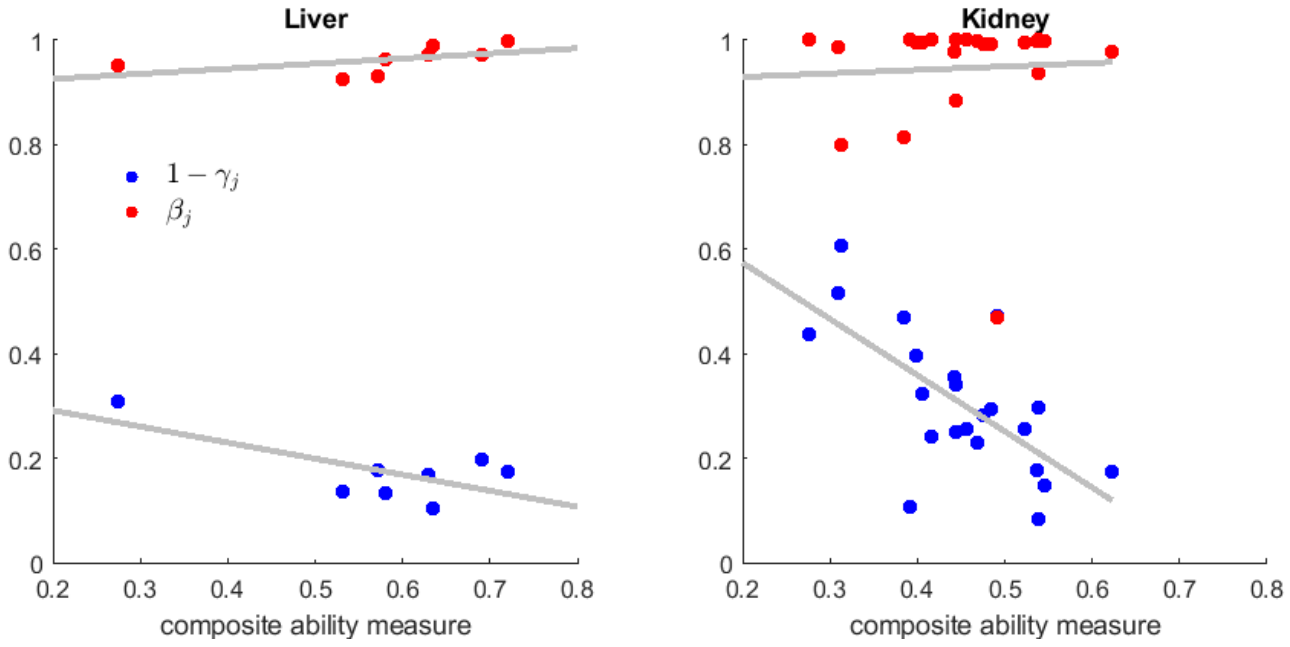
Rearranging terms,

$$p^i_{j1} = (1-\gamma_j) + (1-\pi^i)(\beta_j - (1-\gamma_j)). \tag{15}$$

Given that $(1-\pi^i)$ is increasing in R^i, $\beta_j - (1-\gamma_j)$ is associated with the slope of the estimated $p^i_{j1}-R^i$ relationship reported in Figure 7, while $(1-\gamma_j)$ corresponds to the intercept. Intuitively, higher ability centers (with larger γ_j, β_j) will reject fewer organs with a low risk index, hence the smaller intercept, while simultaneously rejecting more organs with a high risk index, which results in the steeper slope.

At $R^i = \underline{R}$, $\pi^i = 1$ and, hence, $p^i_{j1} = (1-\gamma_j)$ from equation (15). At $R^i = \overline{R}$, $\pi^i = 0$ and, hence, $p^i_{j1} = \beta_j$. Using the $p^i_{j1}-R^i$ relationship that we have estimated for each center, the predicted p^i_{j1} at $R^i = \underline{R}$ provides an estimate of $1-\gamma_j$ and the predicted p^i_{j1} at $R^i = \overline{R}$ provides an estimate of β_j. In practice, we set $\underline{R} = 0.5$ and $\overline{R} = 4$, to be consistent, for both livers and kidneys. This spans

Figure 8: Organ-Quality Specific Ability, by Center



Note: Composite ability measures are based on rejection rates in first position.
Organ-quality specific abilities are derived from the relationship between decisions in first position and organ risk indices.

the range of risk indices that are considered by Collett et al. (2017) and Watson et al. (2012) in their validation analyses.[15] Figure 8 reports the estimated β_j and $1-\gamma_j$ for each center, separately for livers and kidneys. The X axis measures each center's composite ability, q_j, which is the measure we used in the herding tests, based on the rejection rate in first position. Cross-validating the independently constructed ability measures, we see that β_j is increasing and $1-\gamma_j$ is decreasing in q_j. This result is not obtained mechanically because β_j, γ_j, q_j are all derived from decisions in first position. Recall that q_j is measured by the rejection rate in first position, $\overline{p}_1$, for livers and $(1-\overline{p}_1)$ for kidneys.[16]

5.2 Verifying the Assumptions of the Model

We make two assumptions in the model: Assumption 1 specifies that centers are not systematically misinformed; i.e. $\beta_j > 1-\gamma_j$ for all j. As observed in Figure 8, this is evidently the case for each center. Recall that we also assume in the model that the rejection rate in first position, $\overline{p}_1$, is either monotonically increasing or monotonically decreasing in ability. As observed in Figure 8, there is indeed a monotonic relationship between composite ability on the X axis, which is measured by $\overline{p}_1$ for

[15] Given that the $p^i_{j1} - R^i$ relationship has a lower intercept and a steeper slope for higher ability centers, $\underline{R}(\overline{R})$ must be sufficiently low (high) to ensure that the predicted p^i_{j1} is decreasing (increasing) in ability at $\underline{R}(\overline{R})$.

[16] The β_j, γ_j parameters can be linked to the composite ability, q_j, by taking expectations in equation (14):

$$\overline{p}_{j1} = \pi(1-\gamma_j) + (1-\pi)\beta_j.$$

In a superior organ pool, $\pi \to 1$ in the limit, and $q_j \equiv \gamma_j = 1-\overline{p}_{j1}$, as with kidneys. In an inferior organ pool, $\pi \to 0$ in the limit, and $q_j \equiv \beta_j = \overline{p}_{j1}$, as with livers.

livers and $(1-\overline{p}_1)$ for kidneys, and the independently constructed organ-quality specific measures of ability, β_j and γ_j.

Assumption 2 specifies that centers follow their signals in first position; i.e. they accept with a g signal and reject with a b signal. This assumption is satisfied if $\pi_j(g) > \tilde{\pi} > \pi_j(b)$ for all j, where $\pi_j(g)$ is center j's posterior belief that an organ is a G organ after receiving a g signal in first position, $\pi_j(b)$ is the corresponding belief after receiving a b signal, and $\tilde{\pi}$ is the cutoff belief above which centers accept an organ. The ability parameters, β_j, γ_j, that we have estimated tell us how center j's belief that an organ is good responds to b, g signals. To verify Assumption 2 we need, in addition, to estimate the prior belief, π; i.e. the fraction of G organs in the population, and the cut-off belief, $\tilde{\pi}$.

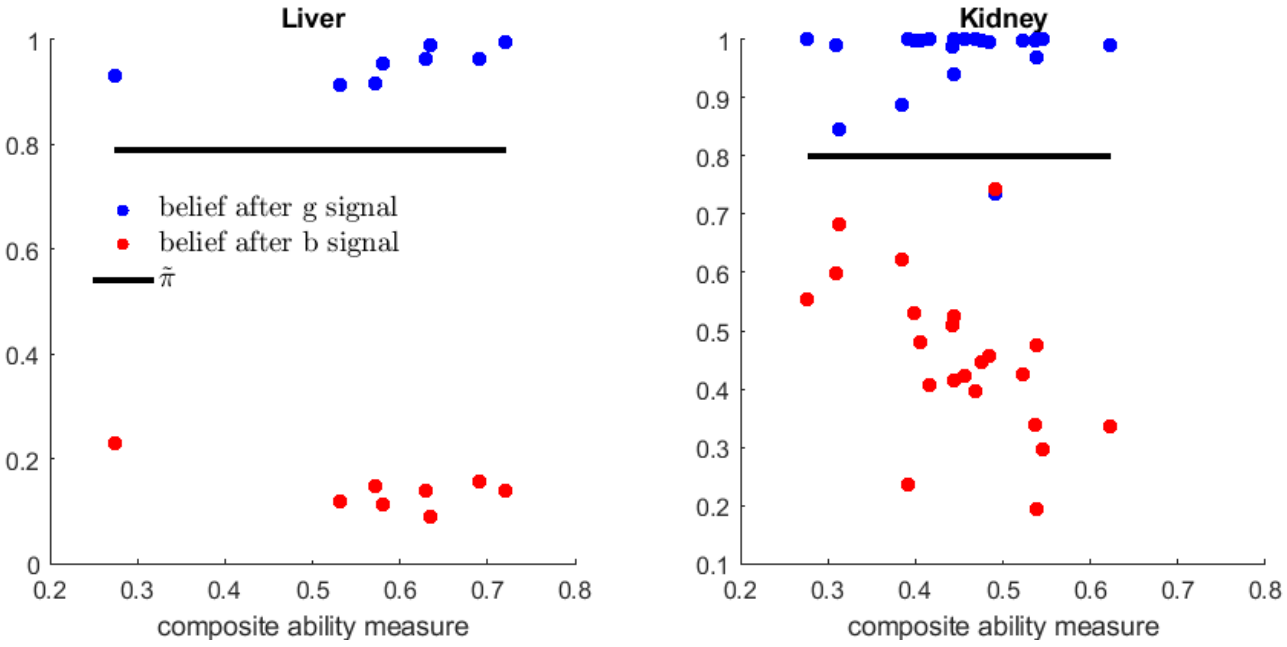
Given the estimated $p^i_{j1}-R^i$ relationship for each center, p^i_{j1} can be predicted for any organ i with risk index R^i. Given the predicted p^i_{j1} and the estimated β_j, γ_j, we can recover $\pi^i = \pi(R^i)$ from equation (14). Although, in principle, π^i corresponding to a given R^i should be the same for all centers, noise in the estimated β_j and γ_j could generate some variation in practice. Our best estimate of π^i is thus the average across all centers that were offered organ i. Averaging the estimated π^i across all organs, we arrive at an estimate of π, the fraction of G organs in the population of organs: 0.48 for livers and 0.74 for kidneys. These estimates allow us to further cross-validate the composite ability measures used in the tests of herding. Recall from Table 2 that the probability of rejection in second position is increasing (decreasing) in $\overline{p}_1$ for livers (kidneys). Based on the model, this implies that livers (kidneys) are drawn from inferior (superior) organ pools and, hence, that $\overline{p}_1$ can be used to measure ability for livers and $(1-\overline{p}_1)$ can be used to measure ability for kidneys. Our estimates of π, based on the risk indices, provide independent support for the inference that livers (kidneys) are drawn from inferior (superior) organ pools.

To verify Assumption 2, all that remains is to estimate $\tilde{\pi}$. All centers follow their signal in first position in the model, which implies that their belief following a g (b) signal lies above (below) $\tilde{\pi}$. While some centers continue to follow their signals in later positions, others will start to herd (i.e. to reject offers regardless of whether they receive a g or a b signal). This is because their beliefs always lie below $\tilde{\pi}$. As $\tilde{\pi}$ increases, the fraction of centers that herd thus increases, with an accompanying increase in the rejection rate (the decisions of centers that follow their signals remain unchanged). To estimate $\tilde{\pi}$ we thus match the overall rejection rate in the data to the rejection rate predicted by the model; there is a unique value of $\tilde{\pi}$ at which the actual and predicted rejection rates match and this will be our best estimate of the $\tilde{\pi}$ parameter.

The simulated method of moments is used to estimate $\tilde{\pi}$. To draw signals for the estimation, we take advantage of the fact that our center-specific probit estimates of the $p^i_{j1}-R_i$ relationship allow us to predict the probability of rejection in first position for any organ-center pair. Since centers always follow their signals in first position, as verified below, this provides us with the probability that the center would receive a b signal in first position and, for that matter, in any position for a given organ. We draw signals in this way, and then predict decisions at each position (given the

previously-estimated values of β_j, γ_j, and π). The average over multiple draws of the signals predicts the overall rejection rate for a given $\tilde{\pi}$, and we then search over all $\tilde{\pi}$ to find the value at which the actual and predicted rejection rates match. Given that β_j, γ_j, π are estimated using decisions at the first position, we match rejection rates at the second position to estimate $\tilde{\pi}$. The $\tilde{\pi}$ estimate, with bootstrapped standard errors in parentheses, is 0.79 (0.004) for livers and 0.80 (0.04) for kidneys. The advantage of estimating $\tilde{\pi}$ at a single position is that we will be able to more stringently validate the model below by comparing the goodness of fit with the data out-of-sample at higher positions.

Figure 9: Posterior Beliefs in First Position, by Center



Note: Composite ability measures are based on rejection rates in first position.
Posterior beliefs are derived after receiving g and b signals, respectively, in first position.
$\tilde{\pi}$ denotes the threshold belief above which centers accept organs.

Having estimated $\tilde{\pi}$, we can now verify Assumption 2. Figure 9 reports the belief after a g signal in first position, $\pi_j(g)$, and the corresponding belief after a b signal, $\pi_j(b)$, for each center, together with $\tilde{\pi}$. We see that $\pi_j(g) > \tilde{\pi} > \pi_j(b)$, with one exception, consistent with Assumption 2. Recall that π is 0.48 for livers and 0.74 for kidneys. Given that $\tilde{\pi}$ is around 0.8, it is quite striking that the posterior belief for livers after a g signal nevertheless satisfies Assumption 2. Notice also that $\pi_j(g)$ is increasing in the composite ability measure, whereas the relationship is reversed for $\pi_j(b)$. This is because all centers start with a common prior belief π in first position, but posterior beliefs (in both directions) respond more to the signals that are received when centers have higher ability.

5.3 Goodness of Fit

Having verified the key assumptions of the model, the next step is to assess its goodness of fit with the data. We estimate $\tilde{\pi}$ by matching rejection rates at second position that are predicted by the model with the data. While we would thus expect a close match in second position, the analysis that follows examines how well the model matches the data at higher positions. As a basis for comparison, we also report rejection rates from an alternative model with no (social) learning. Center ability and the signal-generating process in the alternative no learning model are determined in the same way as in our model, but decisions are now based exclusively on the signals received by each center (without regard to the decisions of preceding centers).

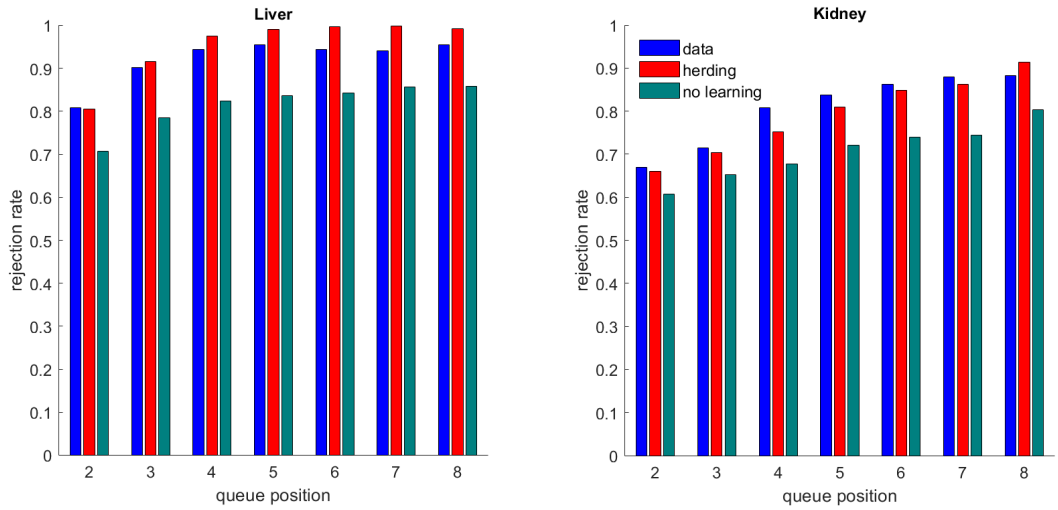
Figure 10 reports rejection rates in the data, predicted by our learning model, and predicted by an alternative no learning model, by position.[17] We see that the rejection rate predicted by the herding model matches closely with the data at each position: the overall prediction error for positions 2-8 is 2.6% for livers and 3.6% for kidneys. The more stringent out-of-sample prediction error for positions 3-8 is not much higher: 3.5% for livers and 4.0% for kidneys. With livers, predicted rejection rates are slightly higher than the data at all positions, whereas with kidneys, they start below and then converge, before overshooting slightly at the highest (eighth) position. This contrasts with the performance of the no learning model, which substantially under-predicts rejection rates at each position. The rejection rate does increase across positions, on account of the fact that lower quality organs travel further down the line on average, but the gap between the predicted and actual rejection rates nevertheless continues to grow across successive positions.

The results in Figure 10 complement the reduced form tests of herding, providing independent evidence that centers are learning from their predecessors' (rejection) decisions. These results can also be used to rule out an alternative interpretation of the observation in Table 1 that transplant success varies little by position. Our interpretation of this finding is that organ deterioration and the patient-organ mismatch, which are both mechanically increasing with position, are less salient in this setting. However, an alternative explanation, which we did not rule out, is that centers account for these factors by becoming more conservative; in the context of our model, this implies that $\tilde{\pi}$ is increasing at higher positions. If this were the case, then given that $\tilde{\pi}$ is estimated at the second position, the model would systematically under-predict rejection rates at higher positions. There is no evidence of such under-prediction, with livers or kidneys, in Figure 10.

While the tests derived from the model indicate that centers are herding, they do not quantify the prevalence of this behavior. Given the estimates of β_j, γ_j, π, $\tilde{\pi}$ and the sequence of centers associated with each organ, we can determine whether a given center at a given position is herding – i.e. that its

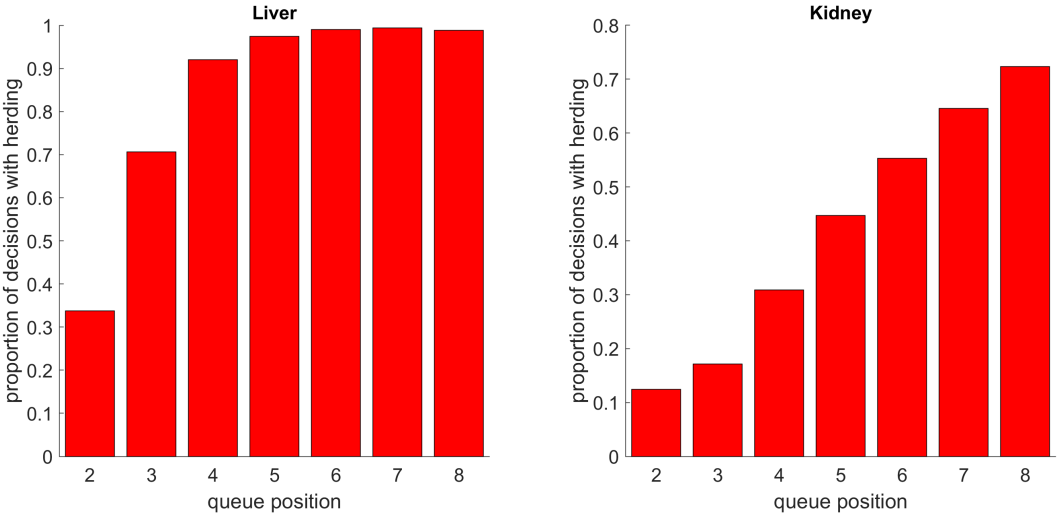
[17] If the model predicts an acceptance but the organ was actually rejected at a given position, then we terminate the sequence and do not make predictions at subsequent positions. If the organ is rejected in the data and the model, we account for the exogenous discards of organs by NHSBT at each position, and the resulting effect on the prior belief of all centers at the next position. The adjustment in beliefs is based on the observed discard rate and the estimated change in the fraction of G organs (based on our estimates of π^i) at each position up to the eighth position. We also make this adjustment when estimating $\tilde{\pi}$ (at the second position).

Figure 10: Goodness of Fit



Note: The model's parameters are estimated at positions 1 and 2.
Rejection rates at positions 3-8 are thus predicted out of sample.

Figure 11: Prevalence of Herding



Note: A center is specified to be herding at a given organ-position if the model predicts it would reject regardless of the signal it received.

prior belief based on preceding decisions is so far below $\tilde{\pi}$ that it will reject regardless of the signal it receives. Figure 11 reports the prevalence of such herding, by position, for livers and kidneys. Herding is very common. For livers, above 30% of centers in second position herd. There is a steep increase in herding at higher positions and, by the sixth position, almost all centers herd. Herding is less prevalent, on average, for kidneys. Nevertheless, over 10% of centers herd in second position and there is a steady increase to 70% by the eighth position.[18]

6 Ability Heterogeneity and Efficiency

The preceding analysis exploits ability heterogeneity to document that centers learn from their predecessors' decisions about organ quality, although information on prior signals is sometimes lost due to herding. The analysis that follows examines the relationship between ability heterogeneity and the prevalence of herding, with its accompanying information loss, which we will see has consequences for the (in)efficiency of organ selection, measured by the fraction of good organs discarded; i.e. false discards and the fraction of bad organs accepted; i.e. false acceptances.

Although the risk index provides an objective measure of organ quality, it does not tell us whether a given organ in the data is good and should be accepted or bad and should be rejected. The efficiency analysis is thus based on 1000 hypothetical "good" livers (kidneys) and 1000 hypothetical "bad" livers (kidneys). Recall that the sequence lengths for the organs in our data rarely exceed eight positions. For each organ we thus draw eight centers randomly from our pool of transplantation centers and then randomly order the selected centers, without regard to their ability, in line with the allocation procedure that is in place. Given the estimated organ-quality and center-specific β_j, γ_j ability parameters and the estimated π, $\tilde{\pi}$, we draw information signals for each center (using β_j for bad organs and γ_j for good organs) and then compute decisions at each position. If an organ is not accepted by the eighth position, then it is assumed to be discarded. In addition, we allow for discards at earlier positions, based on NHSBT's discard rate by position and organ quality, as observed in the data, while simultaneously allowing for the effect of these discards on center beliefs.[19]

Table 6 reports the resulting false discard rates and false acceptance rates, for livers and kidneys. These statistics are reported for two learning models: our herding model, which was validated above, in which the information contained in the signals that are ignored is lost to those that follow and

[18] Based on the model, centers should always reject when they are herding. Centers reject 91% of the time for livers and 81% of the time for kidneys in positions where the estimated model predicts they will herd. In contrast, rejection rates are 67% for livers and 62% for kidneys in positions where centers are not predicted to herd. The discrepancy between the model and observed rejection decisions (at positions where centers are predicted to be herding) can be reconciled by incorporating idiosyncratic shocks, associated with the patient-organ match and organ deterioration, in center decisions. These shocks, which can be interpreted as structural error, wash out when estimating $\tilde{\pi}$ across all decisions, but will be relevant when predicting specific decisions.

[19] Although the efficiency analysis is conducted separately for G and B organs, centers do not know the type of organ they are offered and, hence, beliefs are adjusted to account for discards at each position (both for G and B organs) just as we did earlier when estimating $\tilde{\pi}$ and predicting rejection rates. However, the rates at which organs are discarded are now quality-specific. These rates can be determined at each position, given the observed overall discard rate and the estimated fraction of good organs before and after discarding (based on the risk indices).

a counter-factual pooled information model in which all predecessors' signals are made available to centers when they make their decisions. The results reported in Table 6 can be summarized as follows: (i) Both false discard rates and false acceptance rates are very similar with the two models of learning. Herding does not generate substantial inefficiencies in our setting. (ii) False discards and false acceptances are low overall. This is true when these inefficiencies are measured as rates; i.e. with respect to the number of good and bad organs, respectively, or in absolute terms based on the number of incorrectly assigned organs (over the 2006-2015 analysis period, 13,540 livers and 19,225 kidneys were offered for transplantation).[20] The analysis that follows will examine whether and how ability heterogeneity has contributed to these low levels of inefficiency.

Table 6: False Discards and False Acceptances

model:	herding	pooled information
liver		
false discard rate	0.0306	0.0235
false acceptance rate	0.0357	0.0315
kidney		
false discard rate	0.0362	0.0339
false acceptance rate	0.0358	0.0350

False discard (acceptance) rate is the fraction of good (bad) organs that are discarded (accepted).
These statistics are computed with the estimated ability parameters; i.e. at the observed level of ability heterogeneity.

For the counter-factual analysis that follows, we construct the following measures of ability at different levels of heterogeneity:

$$\hat{\beta}_j(h) = \overline{\beta} + h(\beta_j - \overline{\beta})$$

$$\hat{\gamma}_j(h) = \overline{\gamma} + h(\gamma_j - \overline{\gamma})$$

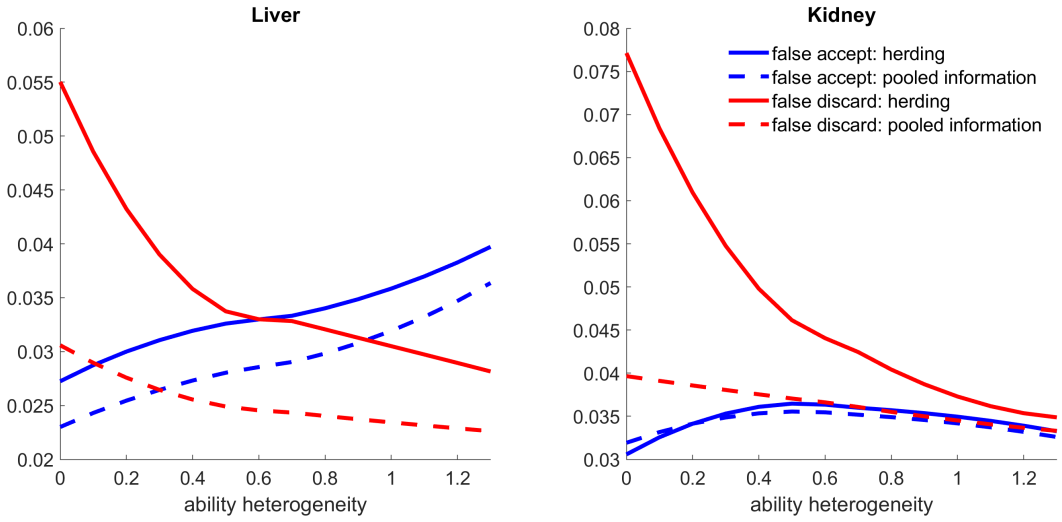
where $\overline{\beta}$, $\overline{\gamma}$ denote the mean abilities across all centers and $h \in [0, 1.2]$ is a heterogeneity parameter. When $h = 0$, all centers have the same ability, $\overline{\beta}$, $\overline{\gamma}$. When $h = 1$, ability levels match what we estimate in the data, β_j, γ_j. In general, there is a mean-preserving increase in ability heterogeneity as h grows larger; centers with higher (lower) than average ability see an increase (decrease) in their ability. We restrict the maximum value of h to 1.2 because the estimated β_j parameters (at $h = 1$) are close to one and further dispersion in ability raises some of these values above one.

Figure 12 plots the false discard rate and the false acceptance rate against ability heterogeneity,

[20] Notice that false acceptance rates and false discard rates are comparable, despite the fact that the β_j parameter, which determines the accuracy of signals received with bad organs, is generally much larger than the γ_j parameter, which determines the accuracy of signals received with good organs, in Figure 8. This is because a single mistake with a bad organ results in a false acceptance, whereas a false discard only occurs when all centers who are offered a good organ reject.

separately for the herding model and the pooled information model and separately for livers and kidneys. The first observation is that false discard rates are declining in ability heterogeneity, whereas false acceptance rates are increasing in heterogeneity for both models. The second observation is that the inefficiency due to herding, measured by the gap between the two models, is decreasing in ability heterogeneity. This is especially true for the false discard rate. We discuss each observation, in turn, below.

Figure 12: False Discard Rate and False Acceptance Rate



Note: False discard (acceptance) rate is the fraction of good (bad) organs that are discarded (accepted). Ability heterogeneity has value one at the level observed in the data.

False discards arise when good organs fail to be accepted by all centers to which they are offered. As heterogeneity increases, higher ability centers are better able to detect good organs, whereas lower ability centers are less able to detect good organs, leaving the average detection ability unchanged. What changes asymmetrically with an increase in heterogeneity is the fraction of decision taken by centers with different levels of ability. When more able centers are positioned earlier, they now detect more good organs and so the less able centers decide less often. When less able centers go earlier, they now pass on more (good) organs to the higher ability centers that follow. Regardless of the ordering of centers, an increase in ability heterogeneity increases the fraction of decisions taken by more able centers, with an accompanying decline in the false discard rate.

False acceptances arise when any center that is offered a bad organ accepts it. Following the same reasoning as above, when more able centers go earlier in line, they will pass on more bad organs to later positions as heterogeneity (and, hence, their ability) increases. Conversely, when less able centers go earlier, they will falsely accept more often as heterogeneity increases (with a commensurate decline in their ability) passing on fewer bad organs to the now more able centers that follow. Regardless of

the ordering, an increase in ability heterogeneity increases the fraction of decisions taken by less able centers, with an accompanying increase in the false acceptance rate.

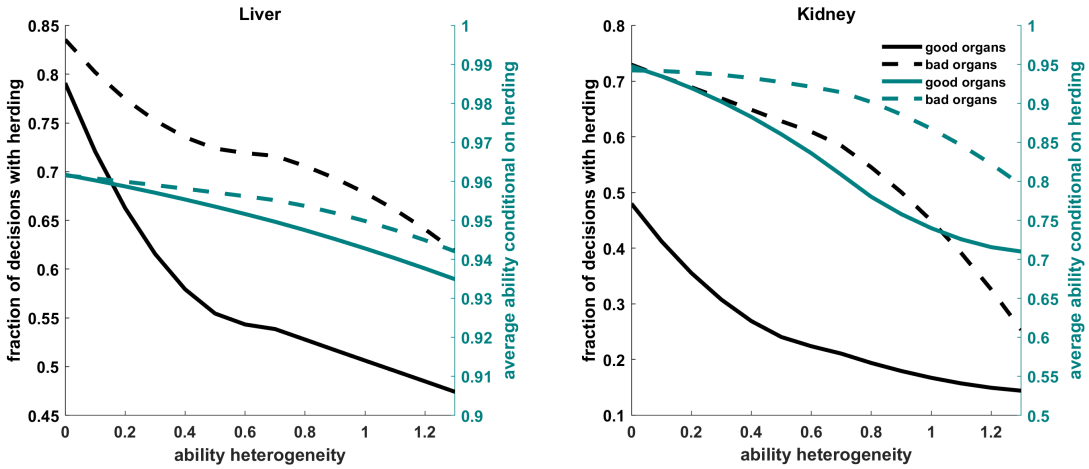
The preceding arguments apply to both the herding model and the pooled information model. To explain why the gap between the two models declines with ability heterogeneity; i.e. the second observation from Figure 12, recall that the only difference between the two models is that signals received by centers that herd are lost to those that follow with the herding model. The information loss and, by extension, the difference in outcomes generated by the alternative learning models will thus depend on (i) the prevalence of herding and (ii) the information loss conditional on herding; i.e. on the ability of centers who herd. In general, the relationship between the prevalence of herding and ability heterogeneity is ambiguous. When higher ability centers follow lower ability centers, herding is less likely and this effect is amplified by an increase in heterogeneity (a mean-preserving spread of the ability distribution). When lower ability centers follow higher ability centers, the converse is true.

Figure 13 plots the relationship between the fraction of decisions with herding and ability heterogeneity, separately for good and bad organs, with livers and kidneys. At each level of ability heterogeneity, h, with its corresponding center abilities, $\hat{\beta}_j$ $\hat{\gamma}_j$, we use the model to predict center decisions and outcomes (acceptances or discards) just as we did in Table 6. We see that the herding rate is declining in ability heterogeneity, without exception. We also see that ability, conditional on herding, is declining in ability heterogeneity. This can be explained by two reinforcing effects: (i) less able centers are more likely to herd and since the prevalence of herding has declined, centers that herd will be drawn from lower in the ability distribution, and (ii) the ability distribution spreads out with an increase in heterogeneity.[21]

Based on Figure 12 and Figure 13, an increase in ability heterogeneity is associated with a decline in the prevalence of herding, the information loss that goes with it and, hence, with inefficiency in organ selection. Although the prevalence of herding is quite high in our data, which corresponds to $h = 1$ in the figures, herding would be much higher if centers were homogeneous in their abilities ($h = 0$). This would substantially increase inefficiencies in organ selection, especially the rate of false discards. While ability heterogeneity reduces herding and its accompanying inefficiencies, the important qualifier to these findings is that they apply to the current setting. In other environments, the inefficiencies due to herding might well be substantial.

[21] Notice that the herding rate is higher with bad organs than with good organs at each level of ability heterogeneity in Figure 13. Although beliefs evolve and centers make decisions just as in the model, we now draw signals for centers that are specific to the quality of the organ; i.e. using $\hat{\beta}_j$ ($\hat{\gamma}_j$) to draw signals with bad (good) organs. Given that $\beta_j > 1-\gamma_j$, as verified above, and hence that $\hat{\beta}_j > 1-\hat{\gamma}_j$, it follows that bad organs are more likely to be rejected than good organs and will thus travel further down the line, where the prevalence of herding is greater on average, as observed in Figure 11. Based on the preceding discussion, this also explains why ability, conditional on herding, is lower for good organs.

Figure 13: Herding Rate and Ability Conditional on Herding



Note: Herding rate is computed as the fraction of decisions in positions 2-8 at which centers are predicted to reject regardless of their signal.
Average ability conditional on herding is based on the composite ability measure used in the herding tests.
Ability heterogeneity has value one at the level observed in the data.

7 Conclusion

There are many economic environments in which prospective buyers, acting sequentially, must choose whether or not to acquire an object. It is often observed that (rejection) decisions across buyers are correlated. One explanation for this correlation is that buyers independently assess that the object is of poor quality. Another explanation is that agents further in line herd behind their predecessors and (rationally) ignore their own assessment of the object's quality. Our research extends the canonical herding model by allowing agents to differ in their ability to assess the quality of the offered object. We develop new tests of herding based on this ability heterogeneity and also examine its efficiency consequences, applied to liver and kidney transplantation in the U.K. over the 2006-2015 period.

Organ transplantation in the U.K. is an ideal setting for our analysis for a number of reasons. First, the ordering of centers for a given organ is independent of their ability to distinguish between good and bad organs. Second, the payoff from transplanting an organ is independent of center ability, as we define it. Third, most decisions are made at very early positions and almost never past the eighth position. This implies that the patient-organ mismatch and organ deterioration, which will mechanically increase by position, are less relevant, as are strategic dynamic considerations in decision-making. Center decisions are based on an assessment of objective organ quality and we detect herding in these decisions with multiple independent tests.

Although herding is common, our analysis indicates that the level of inefficiency, measured by false

discards of good organs and false acceptances of bad organs, is surprisingly low relative to the pooled information benchmark. Although the relationship between herding inefficiency and ability heterogeneity is theoretically ambiguous, in our setting, the prevalence of herding and the accompanying inefficiency are declining in center heterogeneity. In particular, the false discard rate, based on our counter-factual analysis, would have doubled if centers were homogeneous in their abilities. There is, however, no obvious reason why the conditions that reduce herding inefficiencies in the U.K. should hold in other settings, such as the U.S., where organ transplantation is organized very differently and where the ability distribution could well be different. An examination of the efficiency of organ transplantation in such settings, along the lines of the analysis in this paper, together with an investigation of appropriate corrective policies, would thus appear to be fruitful areas for future research.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "When should you adjust standard errors for clustering?," *The Quarterly Journal of Economics*, 2023, *138* (1), 1–35.

Agarwal, Nikhil, Charles Hodgson, and Paulo Somaini, "Choices and Outcomes in Assignment Mechanisms: The Allocation of Deceased Donor Kidneys," Technical Report, National Bureau of Economic Research 2020.

_ , Itai Ashlagi, Michael A Rees, Paulo Somaini, and Daniel Waldinger, "Equilibrium allocations under alternative waitlist designs: Evidence from deceased donor kidneys," *Econometrica*, 2021, *89* (1), 37–76.

Anderson, Lisa R and Charles A Holt, "Information cascades in the laboratory," *The American Economic Review*, 1997, pp. 847–862.

Avery, Christopher and Peter Zemsky, "Multidimensional uncertainty and herd behavior in financial markets," *American Economic Review*, 1998, *88* (4), 724–748.

Banerjee, Abhijit, Emily Breza, Arun G Chandrasekhar, and Markus Mobius, "Naive learning with uninformed agents," *American Economic Review*, 2021, *111* (11), 3540–3574.

Banerjee, Abhijit V, "A simple model of herd behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bohren, J Aislinn, "Informational herding with model misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Cameron, Stephen V and James J Heckman, "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political economy*, 1998, *106* (2), 262–333.

Çelen, Boğaçhan and Shachar Kariv, "Observational learning under imperfect information," *Games and Economic behavior*, 2004, *47* (1), 72–86.

Chamley, Christophe P, *Rational herds: Economic models of social learning*, Cambridge University Press, 2004.

Cipriani, Marco and Antonio Guarino, "Estimating a structural model of herd behavior in financial markets," *American Economic Review*, 2014, *104* (1), 224–51.

Collett, David, Peter J Friend, and Christopher JE Watson, "Factors associated with short- and long-term liver graft survival in the United Kingdom: development of a UK donor liver index," *Transplantation*, 2017, *101* (4), 786–792.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

Dupas, Pascaline, "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica*, 2014, *82* (1), 197–228.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting Others and the Fragility of Social Learning," 2019.

Gale, Douglas, "What have we learned from social learning?," *European Economic Review*, 1996, *40* (3-5), 617–628.

Golub, Benjamin and Matthew O Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–49.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, "Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior," *The American Economic Review*, 1995, pp. 1088–1105.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, "The impact of institutional trading on stock prices," *Journal of Financial Economics*, 1992, *32* (1), 23–43.

Mobius, Markus, Tuan Phan, and Adam Szeidl, "Treasure hunt: Social learning in the field," Technical Report, National Bureau of Economic Research 2015.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Park, Andreas and Hamid Sabourian, "Herding and contrarian behavior in financial markets," *Econometrica*, 2011, *79* (4), 973–1026.

Roth, Alvin E, Tayfun Sönmez, and M Utku Ünver, "Kidney exchange," *The Quarterly Journal of Economics*, 2004, *119* (2), 457–488.

_ , _ , and _ , "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *American Economic Review*, 2007, *97* (3), 828–851.

Watson, Christopher JE, Rachel J Johnson, Rhiannon Birch, Dave Collett, and J Andrew Bradley, "A simplified donor risk index for predicting outcome after deceased donor kidney transplantation," *Transplantation*, 2012, *93* (3), 314–318.

Watson, CJE and JH Dark, "Organ transplantation: historical perspective and current practice," *British Journal of Anaesthesia*, 2012, *108* (suppl_1), i29–i42.

Wermers, Russ, "Mutual fund herding and the impact on stock prices," *the Journal of Finance*, 1999, *54* (2), 581–622.

Zhang, Juanjuan, "The sound of silence: Observational learning in the US kidney market," *Marketing Science*, 2010, *29* (2), 315–335.

8 Appendix (not for publication)

This appendix sets up an alternative model in which centers do not differ in their ability to distinguish between good and bad organs but, instead, are heterogeneous in their cut-off beliefs, $\tilde{\pi}$. First, we show formally (Lemma 1) that there is a positive relationship between a center's cut-off belief $\tilde{\pi}$ and p_1, its probability of rejecting an organ in first position. As noted in Section 4.1, center 1's rejection rate $\overline{p}_1$ is equivalent to p_1, and hence the interpretation of $\overline{p}_1$ in our alternative model is that it proxies for $\tilde{\pi}$. Next, we show (Lemma 2) that there is an unambiguously negative relationship between $\tilde{\pi}$ and p_2, the probability of rejection of the center in second position (conditional on being offered the organ). This is true regardless of whether centers are herding or not. This is inconsistent with the results in Table 2. Finally, we show (Lemma 3) that, without herding, if centers 1 and 2 have the same cut-off beliefs, the effect of an increase in center 1's belief on p_3, the probability of rejection of the center in third position, is the same as an increase in center 2's belief. This is inconsistent with the results in Tables 4 and 5. We formalize our argument within the context of a model with a continuum of signals, for ease of exposition, but the results easily extend to finite signals.

Assume that a G (B) organ generates signals s in $[\underline{s}, \overline{s}] \subset \mathbb{R}$ according to a cumulative distribution function F_G (F_B) with continuous density f_G (f_B). The distribution functions are common to all centers, i.e. centers are homogenous in their abilities to identify G and B organs. We assume further that F_G and F_B satisfy the weak MLRP, in the sense that higher signals are no less likely to be received with good organs than bad organs: $\frac{f_G(s)}{f_B(s)}$ is weakly increasing in s. This assumption is analogous to Assumption 1 in our model, which states that centers are not systematically misinformed. For any $j = 1, 2, ...$, we denote by $\tilde{\pi}_j$ the cut-off belief of center j (i.e. the center placed at position j), by π_j the probability that an organ offered at position j is a G organ (note, that $\pi_1 = \pi$), and by $\pi_j(s_j)$ the posterior belief of center j that, after receiving signal s_j, the organ offered at position j is a G organ. We assume that $\pi_1(\underline{s}) < \tilde{\pi}_1 < \pi_1(\overline{s})$, that is, center 1 rejects (accepts) the organ after the lowest (highest) signal realization. This assumption says that center 1 both accepts and rejects with positive probability and is thus similar to Assumption 2 in our model, which states that each centre, if it is first in line, rejects the organ after a b signal and accepts after a g signal. In what follows (proofs of Claim 2 and Lemma 3), we will also use the following notation: $H(s) := \frac{F_B(s)}{F_G(s)}$.

Lemma 1 p_1 is weakly increasing in $\tilde{\pi}_1$.

Proof: Weak MLRP implies that

$$\pi_1(s_1) = \frac{\pi f_G(s_1)}{\pi f_G(s_1) + (1-\pi) f_B(s_1)} = \frac{1}{1 + \frac{1-\pi}{\pi} \frac{f_B(s_1)}{f_G(s_1)}} \tag{16}$$

is weakly increasing in s_1. Denote by $\tilde{s}_1$ the lowest signal s_1, such that $\pi_1(s_1) = \tilde{\pi}_1$. $\tilde{s}_1 \in [\underline{s}, \overline{s}]$ is uniquely defined and is weakly increasing in $\tilde{\pi}_1$ (because $\pi_1(s_1)$ is weakly increasing in $\tilde{\pi}_1$). Center 1's

optimal decision is to reject upon observing any signal $s_1 < \tilde{s}_1$ and to accept otherwise. Therefore, center 1's rejection probability p_1 is a function of $\tilde{s}_1$ with

$$p_1 = \pi F_G(\tilde{s}_1) + (1-\pi)F_B(\tilde{s}_1)$$

Since $\tilde{s}_1$ is weakly increasing in $\tilde{\pi}_1$, so is p_1. QED

Lemma 2 Center 2's rejection probability p_2 is weakly decreasing in center 1's cut-off belief $\tilde{\pi}_1$, independent of whether there is social learning or not.

Proof:

$$\frac{dp_2}{dp_1} = \frac{dp_2}{d\pi_2} \cdot \frac{d\pi_2}{d\tilde{\pi}_1} \leq 0.$$

The inequality follows from Claims 1 and 2.

Claim 1: π_2 is weakly increasing in $\tilde{\pi}_1$.

Claim 2: p_2 is weakly decreasing in π_2, independent of whether there is social learning or not.

Proof Claim 1: When center 2 is offered an organ it must be that the organ has been turned down by center 1, implying that $s_1 \in [\underline{s}, \tilde{s}_1]$. The probability that this organ is a G organ is therefore

$$\pi_2 = \frac{\pi F_G(\tilde{s}_1)}{\pi F_G(\tilde{s}_1) + (1-\pi)F_B(\tilde{s}_1)} = \frac{1}{1 + \frac{1-\pi}{\pi} H(\tilde{s}_1)} \tag{17}$$

Since $\tilde{s}_1$ is weakly increasing in $\tilde{\pi}_1$, π_2 is weakly increasing in $\tilde{\pi}_1$ if and only if $H(\tilde{s}_1)$ is weakly decreasing in $\tilde{s}_1$. The derivative of $H(\tilde{s}_1)$ with respect to $\tilde{s}_1$ is

$$\frac{f_B(\tilde{s}_1)F_G(\tilde{s}_1) - f_G(\tilde{s}_1)F_B(\tilde{s}_1)}{F_G^2(\tilde{s}_1)},$$

which is weakly negative if and only if

$$f_B(\tilde{s}_1)F_G(\tilde{s}_1) = \int_{\underline{s}}^{\tilde{s}_1} f_B(\tilde{s}_1)f_G(s)ds \leq \int_{\underline{s}}^{\tilde{s}_1} f_G(\tilde{s}_1)f_B(s)ds = f_G(\tilde{s}_1)F_B(\tilde{s}_1).$$

The latter inequality holds because weak MLRP implies that

$$f_B(\tilde{s}_1)f_G(s) \leq f_G(\tilde{s}_1)f_B(s) \text{ for all } s \leq \tilde{s}_1.$$

Therefore, π_2 is weakly increasing in $\tilde{\pi}_1$. QED

Proof Claim 2: Like center 1, center 2 with cut-off belief $\tilde{\pi}_2$ and corresponding signal $\tilde{s}_2$, the lowest signal s_2, such that $\pi_2(s_2) = \tilde{\pi}_2$, will reject the organ if and only if $s_2 < \tilde{s}_2$. Center 2's rejection

probability is therefore

$$p_2 = \pi_2 F_G(\tilde{s}_2) + (1-\pi_2)F_B(\tilde{s}_2). \tag{18}$$

Assume first, that center 2 *does not* learn from center 1's rejection. Then, center 2 behaves as if it was first in line, and so its belief after receiving signal s_2 is

$$\pi_2(s_2) = \frac{\pi f_G(s_2)}{\pi f_G(s_2) + (1-\pi) f_B(s_2)} = \frac{1}{1 + \frac{1-\pi}{\pi}\frac{f_B(s_2)}{f_G(s_2)}} \tag{19}$$

This is independent of π_2, and $\tilde{s}_2$ is therefore also independent of π_2. Taking derivative of (18) with respect to π_2 gives

$$\frac{dp_2}{d\pi_2} = F_G(\tilde{s}_2) - F_B(\tilde{s}_2) \leq 0,$$

because F_G first order stochastically dominates F_B, which follows from weak MLRP.

Assume second, that center 2 *does* learn from center 1's rejection. Then, center 2's prior belief is π_2 and its posterior belief after receiving signal s_2, is

$$\pi_2(s_2) = \frac{\pi_2 f_G(s_2)}{\pi_2 f_G(s_2) + (1-\pi_2) f_B(s_2)} = \frac{1}{1 + \frac{1-\pi_2}{\pi_2}\frac{f_B(s_2)}{f_G(s_2)}} \tag{20}$$

It is easy to see that $\pi_2(s_2)$ is increasing in π_2, and, therefore, $\tilde{s}_2$ is decreasing in π_2.

Taking derivative of (18) with respect to π_2, we obtain

$$\frac{dp_2}{d\pi_2} = F_G(\tilde{s}_2) - F_B(\tilde{s}_2) + \pi_2 f_G(\tilde{s}_2)\frac{d\tilde{s}_2}{d\pi_2} + (1-\pi_2) f_B(\tilde{s}_2)\frac{d\tilde{s}_2}{d\pi_2},$$

which, based on our previous argument and the fact that $\frac{d\tilde{s}_2}{d\pi_2} < 0$, is no greater than zero. QED

Lemma 3: Suppose there is no herding. Then, the effect of an increase in center 1's cut-off belief, $\tilde{\pi}_1$, on center 3's rejection probability, p_3, is of the same magnitude as the effect of an increase in center 2's cut-off belief, $\tilde{\pi}_2$, i.e.

$$\frac{\partial p_3}{\partial \widetilde{\pi}_1} = \frac{\partial p_3}{\partial \widetilde{\pi}_2}.$$

Proof: The proof procedes in 3 steps. We first show that $\widetilde{\pi}_1 = \widetilde{\pi}_2$ implies

$$\tilde{s}_1 = \tilde{s}_2 \text{ and } \frac{d\tilde{s}_1}{d\widetilde{\pi}_1} = \frac{d\tilde{s}_2}{d\widetilde{\pi}_2}. \tag{21}$$

(21) follows because, if center 2 does not learn from center 1's rejection, $\pi_1(s_1)$ is identical to $\pi_2(s_2)$ (compare (16) and (19)). Therefore, if $\widetilde{\pi}_1 = \widetilde{\pi}_2$, it must be that $\tilde{s}_1$ (its derivative with respect to $\widetilde{\pi}_1$) is identical to $\tilde{s}_2$ (its derivative with respect to $\widetilde{\pi}_2$).

We next show

$$\frac{\partial \pi_3}{\partial \widetilde{\pi}_1} = \frac{\partial \pi_3}{\partial \widetilde{\pi}_2}. \tag{22}$$

When center 3 is offered an organ it must be that the organ has been turned down by both centers 1 and 2, implying that $s_j \in [\underline{s}, \tilde{s}_j]$, $j = 1, 2$. The probability that this organ is a G organ at round 3 is therefore

$$\pi_3 = \frac{\pi_2 F_G(\tilde{s}_2)}{\pi_2 F_G(\tilde{s}_2) + (1-\pi_2) F_B(\tilde{s}_2)} = \frac{1}{1 + \frac{1-\pi_2}{\pi_2} H(\tilde{s}_2)}. \tag{23}$$

Using (17), we can replace π_2 in (23) to obtain

$$\pi_3 = \frac{1}{1 + \frac{1-\pi}{\pi} H(\tilde{s}_1) H(\tilde{s}_2)}.$$

This expression is symmetric in $\tilde{s}_1$ and $\tilde{s}_2$, and so (21) implies (22).

Finally, we show

$$\frac{\partial p_3}{\partial \widetilde{\pi}_1} = \frac{\partial p_3}{\partial \widetilde{\pi}_2}. \tag{24}$$

Like centers 1 and 2, center 3 does not learn from its predecessors and, after receiving signal s_3, has belief

$$\pi_3(s_3) = \frac{\pi f_G(s_3)}{\pi f_G(s_3) + (1-\pi) f_B(s_3)} = \frac{1}{1 + \frac{1-\pi}{\pi} \frac{f_B(s_3)}{f_G(s_3)}}.$$

Assume that center 3 has a cut-off belief $\tilde{\pi}_3$ and define the corresponding signal $\tilde{s}_3$ as the lowest signal s_3, such that $\pi_3(s_3) = \tilde{\pi}_3$. Given the expression for $\pi_3(s_3)$, $\tilde{s}_3$ is independent of $\widetilde{\pi}_i$ (and $\tilde{s}_i$) for $i = 1, 2$. Center 3's rejection probability is

$$p_3 = \pi_3 F_G(\tilde{s}_3) + (1-\pi_3) F_B(\tilde{s}_3). \tag{25}$$

The derivative of p_3 with respect to $\widetilde{\pi}_j$, $j = 1, 2$, is

$$\frac{dp_3}{d\widetilde{\pi}_j} = \frac{\partial \pi_3}{\partial \widetilde{\pi}_j} [F_G(\tilde{s}_3) - F_B(\tilde{s}_3)],$$

and so (24) follows from (22). QED