Evolutionary Ecology Research, 2000, **2**: 493–508

Towards a theory of mutual mate choice: Lessons from two-sided matching

Carl T. Bergstrom and Leslie A. Real*

Department of Biology, Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA

ABSTRACT

Theoretical models of mate choice and sexual selection typically make one of two simplifying assumptions. Either mate preferences are assumed to be uniform (e.g. all females have the same preferences with respect to males), or mate choice is assumed to be a one-sided affair (e.g. females do all the choosing). Recent empirical studies suggest that, in many cases, neither assumption holds. In this paper, we show how two-sided matching – a branch of game theory developed in the economics literature – can be used to model mutual mate choice with non-uniform mate preferences. The economics literature is reviewed, and a number of biological applications are suggested. We characterize a systematic conflict of interest between males and females over the optimal matchings in mutual mate choice systems. Moreover, we observe that the component of choice that confers the most benefit in this conflict is not choice in the conventional sense of accepting or rejecting courtships, but instead the power to choose the individuals to whom one displays.

Keywords: assortative mating, coalitions, game theory, group formation, mating systems, sexual selection.

INTRODUCTION: A ONE-SIDED VIEW OF MATE CHOICE

Through the decision-making processes involved in mate choice, individual behavioural mechanisms generate the evolutionary dynamics of sexual selection. Over the past two decades, recognition that sexual selection is fundamentally a behavioural process has led to the formulation of new classes of models that explicitly incorporate individual decision-making (Janetos, 1980; Parker, 1983; Real, 1990, 1991; Drombrovsky and Perrin, 1994; Getty, 1995; Wiegmann *et al.*, 1996). These models aim to account for patterns of mating within populations (and the resultant direction of sexual selection) as a consequence of individual decisions.

In this paper, we address two of the most common assumptions underlying individual choice models of sexual selection. In most mate choice models, one sex is treated as passive (but see Parker, 1983, discussed below). That is, active choice is usually restricted to a single sex. Since Bateman (1948), theories of mate choice have generally assumed that the female

* Author to whom all correspondence should be addressed. e-mail: lreal@biology.emory.edu

would be the choosy sex, since females are limited by resources for producing young while males are only limited by access to potential mates. However, in some systems, males appear to be the choosy sex (e.g. Manning, 1980). Regardless of which sex does the choosing in a model where one sex is passive, the choosy sex makes decisions based upon the distribution of fitness gains from potential mates balanced against the costs of competition and search (Real, 1990; Wiegmann *et al.*, 1996).

In some biological cases, the assumption that only one sex actively selects among mates is reasonable. For example, in the mite species *Tetranychus urticae*, males guard duetonymph females and prefer those females that are closest to eclosion (Everson and Addicott, 1982). The females obviously have no potential role in choice here, since they are still in their quiescent form.

However, mounting empirical evidence suggests that few systems are as simple as these mites. Mate choice is seldom structured so that one side is entirely passive (Cunningham and Birkhead, 1998). Instead, males and females are actively engaged in mutual choice where each can potentially accept or reject the other. For example, crested auklets, *Aethia cristatella*, engage in a mutual display of crest-ornaments during courtship. When Jones and Hunter (1993) placed model auklets within the breeding colony, both males and females directed their courtship displays to the models with the largest crests, suggesting mutual mate selection. Recently, Kraak and Bakker (1998) have shown mutual mate choice in sticklebacks, as discussed further below.

Mutual mate choice may be much more common than we think, and mate choice models must be constructed that incorporate this important feature of sexual selection (Johnstone *et al.*, 1996). Using a game-theoretic approach, Parker (1983) was the first to outline theories of mate choice that explicitly incorporate mutual choice. Parker's models assume that all potential mates are accessible at all times and there is no search among mates. Real (1991) incorporated mutual choice into a sequential choice model that incorporated searching behaviour. Parker's model and Real's model lead to similar conclusions and generate some very important common features of mating systems that do not arise in single-sex choice models, most notably assortative mating. In this paper, we expand on Parker's mutual choice approach, relaxing his assumption – discussed below – of uniform preferences.

Current modelling efforts usually assume that all individuals of a given sex have the same preferences; that is, that preferences are uniform (Parker, 1983; McNamara and Collins, 1990; Johnstone, 1997), or at most homotypic, such that individuals prefer mates similar to themselves (Alpern and Reyniers, 1999). Such assumptions may be valid for sticklebacks. Bright red males are choosy, and prefer the same large fecund females; similarly, females share a common preference for the brightest-coloured males (Kraak and Bakker, 1998). Nonetheless, in many systems we know that individuals vary considerably in their mating preferences (Jennions and Petrie, 1997; Bakker *et al.*, 1999; Widemo and Sæther, 1999).

Variation in mating preferences may be a consequence of variation in the experience and condition of individuals within the population. For example, recent evidence suggests that the mating preferences of the wandering albatross, *Diomedea exulans*, are correlated with age (Jouventin *et al.*, 1999). First-time breeders prefer mates of similar age, and widowed birds prefer to mate with birds that have also lost mates. Real (1991) incorporated individual variation in mate choice arising from a different source; in his model, variation in choice is a consequence of varying costs across individuals, despite a preference structure that is assumed to be the same across individuals. Genetics may also play a role; in recent years, a number of studies have investigated the importance of the major histocompatibility complex (MHC) genes in determining the mating preferences of several mammalian species, including humans. There is good evidence, at least in house mice, of negative assortative mating by MHC type (see the review by Penn and Potts, 1999). Examples like these indicate that variation in mating preferences may be a ubiquitous outcome of both genetic and environmental heterogeneity.

Thus we are at a critical juncture in the formation of mate choice theory. Traditional theory has focused primarily on individual choice behaviour, even though we suspect that mutual mate choice is common. Traditional theory has generally ignored the consequences of individual variation in mating preferences, even though we know there are both genetic and environmental conditions that promote polymorphism in preferences. Consequently, we need to develop a more comprehensive theory that incorporates mutual choice with individual variation in preferences. In economics, an extensive literature has been developed to treat structurally similar problems. We believe that the solutions developed there will also have important biological implications. In this paper, we take the economic theory of two-sided matching and apply it to the biological problem of mutual mate choice.

AN ALTERNATIVE APPROACH: TWO-SIDED MATCHING THEORY

Often, in economic systems, agents form pairs or groups: employers and employees, buyers and sellers, schools and students. When both sides have preferences over the pairings formed, predicting the outcome of the matching process becomes non-trivial. Which pairings will occur? Which pairings will be stable? Which matching processes will generate stable matchings? What sorts of strategic behaviour should individuals employ to achieve their desired pairings? To answer these questions, game theorists and economists have developed a theory of two-sided matching, in which agents are assigned preferences over pairings or groupings, and the dynamics and stability of pair or group formation are explored.

The two-sided matching problem was first formulated by Gale and Shapley (1962). They presented a simple model in which college applicants have ordinal preferences over schools, and colleges have ordinal preferences over applicants. How, given these preferences, could college applicants be matched to schools so as to leave both the students and the colleges as satisfied as possible? The authors derived a clever algorithm (described below) designed to create efficient pairings. Their algorithm matches students and schools in such a way that no student wishes to leave her current school for an alternative institution that would be willing to admit her. Subsequent authors expanded upon Gale and Shapley's work, extending their theoretical framework while applying two-sided matching theory to problems ranging from labour markets (Shapley and Shubik, 1972; Crawford and Knoer, 1981; Kelso and Crawford, 1982; Roth, 1982, 1984, 1990, 1991) to human courtship and marriage (Knuth, 1976; Becker, 1981; Bergstrom and Lam, 1989; Bergstrom and Bagnoli, 1993) to the annual ritual of sorority rush (Mongell and Roth, 1991).

In a similar fashion, males and females during courtship and mating may form different preferences based on the biological characteristics of the individuals. Female 1 may prefer male 1 to male 2 or male 3, while female 2 prefers male 2 to male 1 or male 3. The interesting dynamics arise because the males also differ among themselves in their preferences over females. How are stable pairings established given this individual variation in mating preferences?

Terminology and definitions

To introduce this area of game theory, a number of definitions will be needed. Here, and in many of the subsequent examples, we follow (closely, although in less formal fashion) both the general notation and the basic line of presentation employed by Roth and Sotomayor (1990) in their definitive monograph *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis*. For simplicity, we focus on the special case of pair formation from two disjoint classes. We can envision these classes as the set of all males in a population, and the set of all females in that population. Similarly, we can envision the pairings as mating partners. For this reason, matching systems of this type are often called 'marriage markets' in the economics literature. More complicated matching systems, some of which are considered below, can be defined similarly.

Consider a population, each member of which falls into one of two sets: the set of all males $M = \{m_1, m_2, m_3, \ldots, m_j\}$ and the set of all females $F = \{f_1, f_2, f_3, \ldots, f_k\}$. Let each individual m_i or f_i have a list of strict pairing preferences P over the individuals in the other set. For example, a female f_i might have preferences $P(f_i) = m_1, m_4, f_i, m_3, m_2$, meaning that male m_1 would be her first choice, m_4 would be her second choice, and she would rather remain 'single' (represented by pairing with herself, f_i) than form a pair with either m_2 or m_3. The preference of female f_i for male m_1 over m_4 is expressed by the $>_{f_i}$ symbol: $m_1 >_{f_i} m_4$. Male m_h is said to be *acceptable* to female f_i if she prefers pairing with m_h to remaining single – that is, if $m_h >_{f_i} f_i$.

A matching

$$\mu = \begin{Bmatrix} m_1 & m_2 & m_3 & \ldots \\ f_4 & f_2 & f_1 & \ldots \end{Bmatrix}$$

is simply a list of all the pairings in the population (where having oneself for a mate means that one remains single). We indicate the mate of an individual x under matching μ by using $\mu(x)$ for short.

Now we are ready to consider the notion of the stability of a matching. An individual is said to *block* the matching μ if he or she prefers remaining single to taking the mate assigned by μ. A pair m and f are said to block the matching μ if they are not matched by μ, but prefer one another to their mates as assigned by matching μ. In the above notation, (m, f) is a blocking pair if, and only if, $f >_m \mu(m)$ and $m >_f \mu(f)$. Put another way, given matching μ, a blocking pair is a pair that would willingly abandon their mates as determined by μ and elope instead with one another. Finally, the matching μ is defined as *stable* if it is not blocked by any individual or pair (Gale and Shapley, 1962; Roth and Sotomayor, 1990). Much of two-sided matching theory is concerned with determining the conditions under which stable matchings exist, and by what matching algorithms these matchings can be achieved.

Readers familiar with economic theory may at this point be curious as to the relationship between the concept of Pareto optimality and the stability of a matching. Pareto optimality requires that no change exists that betters (or leaves equally well-off) every individual in the population. The concept of a stable matching is stronger than that of a 'Pareto optimal' matching, in that every stable matching is Pareto optimal, but not every Pareto optimal matching is stable. Pareto optimality requires that no two individuals wish to elope together *and* would receive the consent of their partners. Stable matching, by contrast, requires that no two individuals wish to elope together, whether or not their partners would consent.

An example

It may be helpful to look at a concrete example. Consider a population consisting of three females (Ann, Betty and Carol) and three males (Dave, Ed and Frank). The preferences of each individual are given in Table 1.

In this matching system, there are two stable matchings. One (call it μ_1) pairs Betty with Dave, Ann with Ed, and Carol with Frank. The other, μ_2, pairs Betty with Dave, Ann with Frank, and Carol with Ed. Any other matching will allow at least one blocking individual or pair. For example, the matching that pairs Betty with Frank, Ann with Dave, and Carol with Ed has as the blocking pair Betty and Dave; in this matching, Betty and Dave would willingly elope together. In fact, since Betty and Dave are one another's first choices, *any* matching μ' which does not pair them together will be blocked by this pair. If we are interested in finding all stable matchings, we can remove Betty and Dave from the preference list, yielding the reduced preference list given in Table 2.

In this very simple system, an interesting sort of conflict emerges. In mate selection models, it is typically assumed that individuals of the same sex – competing with one another for mates – have conflicting interests among themselves. In this two-sided matching example, by contrast, the conflict of interest crosses sex lines. Ed and Frank share a common interest, and together they find themselves in conflict with Ann and Carol. Ed and Frank would both like to achieve stable matching μ_1, while Ann and Carol would both like to achieve stable matching μ_2. Neither allows eloping pairs, but in μ_1 the males get their first choices (from the reduced preference list) and the females get their second choices, whereas in μ_2 the females get their first choices and the males get their second choices.

Table 1. Individuals and their pairing preferences: the full system

Individual	First choice	Second choice	Third choice
Ann	Dave	Frank	Ed
Betty	Dave	Ed	Frank
Carol	Ed	Frank	Single
Dave	Betty	Carol	Ann
Ed	Betty	Ann	Carol
Frank	Carol	Ann	Betty

Table 2. Individuals and their pairing preferences: the reduced system, with Dave and Betty – always one another's first choices – removed (see text)

Individual	First choice	Second choice
Ann	Frank	Ed
Carol	Ed	Frank
Ed	Ann	Carol
Frank	Carol	Ann

This conflict of interest between the sexes (and common interest among members of the same sex) is not merely a quirk of this particular example, but instead a general feature of the two-sided matching models with two separate sexes and monogamous pairings. Which stable matching occurs in a given biological system will depend on which sex dominates in the choice structure. We will see this more clearly later when we consider the algorithms by which stable matchings are achieved.

Two-sided matching under uniform preferences

When all preferences are uniform – that is, when all males have the same preferences over females and vice versa – it is easy to see that a unique stable matching exists. To see this for monogamous mating systems, label the members of each sex by the preferences of the other sex (so that the first-ranking male m_1 is the first choice of the females, m_2 is the second choice, etc.). Under this system, the only possible stable matching will be

$$\begin{Bmatrix} m_1 & m_2 & m_3 & m_4 & \dots \\ f_1 & f_2 & f_3 & f_4 & \dots \end{Bmatrix}$$

Any other matching will have at least one pair composed of males and females with different ranks. When (m_i, f_j) is the lowest-numbered of these pairs, a blocking pair can be found in either (m_i, f_i) (when $i < j$) or (m_j, f_j) (when $j < i$). Similar arguments hold for alternative matching systems.

For monogamous mating systems, a unique stable matching exists even when only one sex has uniform preferences. We can prove this by describing a *matching procedure* – a set of rules by which a matching is generated – that always yields a stable matching. Suppose that the male sex is the sex with uniform preferences. We can order the females by the males' preferences, from most desirable f_1 to the least desirable f_k. Create a matching $\hat{\mu}$ as follows: f_1 gets her first choice of males, f_2 gets her first choice from among the remaining males, and so on. The matching $\hat{\mu}$ is clearly stable; no blocking pairs exist because every female has her favourite male from among the set of all males who are not paired to a female more desirable than her; every male has the most desirable female that will accept him. Moreover, we can easily prove that any other matching $\mu \neq \hat{\mu}$ is unstable. Let f_x be the first female who does not get the mate she would have had in matching $\hat{\mu}$. She will be paired by μ to a male worse than she got by $\hat{\mu}$, and so would be willing to elope with $\hat{\mu}(f_x)$, her mate under matching $\hat{\mu}$. Because all females more desirable than f_x have the same mate they had in $\hat{\mu}$, male $\hat{\mu}(f_x)$ will have a mate less desirable than at $\hat{\mu}$. He will be willing to elope with f_x and, therefore, $(f_x, \hat{\mu}(f_x))$ is a blocking pair for matching $\mu \neq \hat{\mu}$.

Gale and Shapley's deferred acceptance algorithm

What happens when preferences are not uniform? One of the most remarkable results from two-sided matching theory is that, even under non-uniform preferences, a stable matching (or set of stable matchings) exists in *every* monogamous matching system. To prove this, it is sufficient to describe an algorithm by which a stable matching can be constructed for any such system. We need not suppose that pairing actually occurs by this algorithm in the system we are considering. Rather, the algorithm simply serves as a tool in the proof that a stable matching exists. Below, we outline the *deferred acceptance algorithm*, originally presented by Gale and Shapely (1962).

We designate one sex – male, in the description that follows – as the 'courting' sex. Each male *m* displays to his first-choice female *f*; each female who has received one or more displays rejects all but her favourite male, and keeps her favourite male 'engaged' for the time being. The matching procedure then proceeds repeatedly through the following steps.

1. Each male not currently engaged displays to his favourite female that has not already rejected him. If no acceptable females remain, he remains unmated.
2. Each female who has received one or more courtship displays in this round rejects all but her highest-ranked acceptable male. This may involve rejecting a previously engaged male.

After a finite number of rounds, no new displays will be made (Gale and Shapley, 1962). At this point, the algorithm terminates. All females are paired with the male to whom they are currently engaged; individuals not engaged remain unmated. The matching μ generated in this way is easily seen to be stable. No male wishes to leave his mate at μ for a female who prefers him to her mate at μ, because each male reached his current mate by sequentially courting females in order of preference. No female wishes to leave her mate at μ for a male who prefers her to his mate at μ, because she will have already received a courtship display from any male who is not matched to a female that he prefers to her.

Reversing the algorithm, so that the females display and the males accept or reject courtships, will also lead to a stable matching; this matching may be a different one than that found by the male-courtship form of the algorithm. However, the set of individuals remaining unmated is the same in every stable matching of any given monogamous mating system (McVitie and Wilson, 1970).

APPLICATIONS

In this section, we consider three interesting avenues of inquiry arising from two-sided matching theory. Our aim is not to provide a comprehensive treatment of the theory, but rather to convey the general flavour of the two-sided matching approach, to illustrate the power of this approach and to show how this theory can generate a number of interesting and counterintuitive predictions.

Is monogamy unusually stable?

Comparative analysis of alternative mating systems has long been a topic of considerable interest in evolutionary ecology (Andersson, 1994). Not only do mating systems differ in organization and individual tactics, they might also differ in underlying stability. As demonstrated in the previous section, at least one stable matching exists for every monogamous pairing system. What about polygamous systems?

Gale and Shapley (1962) present an illustration analogous to polygamous mating: the matching of students to colleges. They find that in polygamous systems a stable matching will always exist, given two rather strict conditions: (1) each male must be paired with a pre-determined number of females (or vice versa in polyandrous systems), although this number can be different from male to male; and (2) interactions among females do not affect preferences, in that each female is concerned only with the identity of the male with whom she will be mated, and not with the identities of the other females also matched to that male; similarly, each male has preferences only over the individual identities of the females, and not over pairs or triplets, etc., of females. An easy way to see that such systems will always have a stable matching was suggested by Knuth (1976). Where each male m_i takes $n(m_i)$ females, replace male m_i with $n(m_i)$ proxy males each sharing the preferences of male m_i. Applying the deferred acceptance algorithm for monogamous matching to this new population, a stable polygamous matching is generated.

Conditions (1) and (2) above are reasonable for Gale and Shapley's student–college matching example, and may be adequate for some polygamous mating systems as well. In blackbirds, many females pair with a single male, but all act as independent pairs (Beletsky and Orians, 1996). More often, however, we suspect that these conditions will not be met in biological mating systems. For example, the number of mates with whom a given male pairs may depend on the identities and fertilities of those mates, violating the first condition. When the number of mates taken by a given male is variable, a female may prefer to be the sole mate of an otherwise less desirable male rather than to be the secondary mate of an otherwise more desirable male, violating the second condition as well. In other biological systems, polygamous sets are commonly related (e.g. siblings), again violating the second condition.

When conditions (1) and (2) are relaxed, stable matching need not exist in a polygamous system. In the Appendix, we present two examples of polygamous systems that do not allow *any* stable matching. In the first example, males have preferences over pairs of females rather than merely over individual females. In the second, females are concerned with the identities of the other females with whom they will share their mate.

Of course, not all matching in biological systems takes place in the context of mate choice. Among mammals, for example, members of the same sex often form pairs, trios or larger groups to forage, hunt, deter predation or defend resources (Packer *et al.*, 1991; Gompper *et al.*, 1997; Sterck *et al.*, 1997; Waterman, 1997; Watts, 1998). We can apply matching theory to these systems as well. Coalition formation systems such as these do not necessarily allow the existence of a stable matching. Gale and Shapley (1962) again provide an illustration. Consider four males – three adults A, B, C, and an adolescent D – seeking to form mate-guarding coalitions. Male A prefers to ally himself with B, male B prefers to ally himself with C, and male C prefers to ally himself with A. Each individual ranks the youngster D as its last choice. No matching will be stable, because whoever of A, B or C is paired with D will want to leave D, and one of the other two will be willing to form a blocking pair. Similarly, three-way matching systems, in which individuals from three separate classes (e.g. reproductive male, reproductive female and helper-at-the-nest) are matched into trios, need not allow a stable matching (Alkan, 1986; Roth and Sotomayor, 1990).

Thus monogamous mating systems are indeed unusually stable, in the sense that, under perfect information, they always allow at least one stable matching. Many other systems, including polygamy and general coalition formation, may allow *no* stable matchings. This generates testable predictions for the question of whether serially monogamous systems will have different rates of partner turnover than do polygamous systems. In the limit of perfect information and constant preferences, we would expect continual break-up and remating in polygamous systems, because no stable matching exists. In contrast, we might expect stasis in a monogamous system, which must allow at least one stable matching.

However, the existence of a stable matching does not automatically imply that stasis will be reached, even under perfect information. Although every monogamous mating system must have a stable matching, the breakdown of unstable matchings by 'eloping pairs' does not always lead to a stable matching. Instead, certain sequences of divorce and eloping simply progress through a cycle of unstable matchings, returning to the starting point (Knuth, 1976). Still, from any unstable matching in a monogamous mating system, there exists some sequence of divorces and elopings which leads to a stable matching (Roth and Vande Vate, 1990; but see Tamura, 1993).

Conflict of interest between the sexes

In the example on pp. 497–498, we noted a conflict of interest between the two sexes, and common interests among the members of each sex, with regard to the preferred stable matching. Here, we show that this was not a coincidence, but that instead such cross-sex conflicts and within-sex common interests are general properties of monogamous two-sided matching models. This is somewhat surprising, since males are seemingly in competition with one another for females and vice versa.

When all males like matching μ at least as much as matching μ', we say (following the terminology of Roth and Sotomayor, 1990) that $\mu \geq_M \mu'$. When, in addition, at least one male strictly prefers μ, we write $\mu >_M \mu'$. Similarly, where female preferences are concerned, we write $\mu \geq_F \mu'$ or $\mu >_F \mu'$. A stable matching μ is defined as *M-optimal* if no male prefers any other stable matching μ' (parenthetically, note that some males may prefer certain unstable matchings to even the M-optimal stable matching). *F-optimality* is defined analogously. A mate is said to be *attainable* by an individual if the individual is paired with that mate in some stable matching; an M-optimal matching pairs each male with his favourite attainable mate, and an F-optimal matching pairs each female with her favourite attainable mate.

Remarkably, despite the conflicting interests among males and among females when non-stable matings are considered, in monogamous mating systems an M-optimal and an F-optimal stable matching always exist (Gale and Shapley, 1962; Roth and Sotomayor, 1990). This means that, of all the possible stable matchings, *all* males are able to agree unanimously on a consensus favourite stable matching. Similarly, all females are able to agree on a consensus favourite, although it will generally not be the matching preferred by the males. In fact, as we show below, they are typically antagonistic.

These authors also demonstrate that the male-courtship deferred acceptance algorithm generates the M-optimal mating (and the female-courtship version generates the F-optimal mating). In the male-courtship deferred acceptance algorithm, each male displays to females in order of preference and one can prove that no male is ever rejected by an attainable female; therefore, each male is paired with his most-desirable *attainable* female. (Again, note that under this algorithm males will be rejected by any non-attainable females to whom they propose and therefore may not be paired with their most-desirable female overall.) While intuitively it might seem that females would be getting the best of the situation, in practice, precisely the converse holds. Under this system, the optimal matching for the males as a group – the M-optimal matching – would be generated. The implication of this result is striking: *The component of choice conferring the major benefit is the power to choose the individuals to whom one displays, not the power to choose whether to accept or reject courtships*, or even the power to keep suitors 'engaged' until a better alternative comes along.

This result has intriguing empirical implications. We often assume that, because we see females rejecting mates, female choice is dominating and females are determining the mating structure of the population. Indeed, observing a male-courtship deferred acceptance procedure, we would be prone to conclude that this is a ‘female choice’ system, because females have the power to accept or reject males, and even to keep males on hold, stringing them along only to be later rejected in favour of preferable suitors. In mutual choice situations, however, the greatest power may lie in the males’ courtship behaviour, and the stable mating structure might be determined mostly by what males are choosing to do in terms of display.

Moreover, the M-optimal matching is not only the best stable matching for the males, it is also always the worst stable matching for the females (Knuth, 1976). In fact, male and female preferences conflict in this way over *any* pair of stable matchings, not just the M- and F-optimal ones. Given μ and μ' are two stable matchings, $\mu >_M \mu'$ implies that $\mu' >_W \mu$ and vice versa (Roth and Sotomayor, 1990). Therefore, while all males reach consensus regarding the best stable matching, as do all females, the sexes inevitably oppose one another in their preferences across alternative stable matchings.

We can also view the conflict at the level of individual spouses. Knuth (1976) proves the following theorem. Consider two individuals m_1 and f_1, paired together under stable matching μ but paired with f_a and m_a respectively under some other stable matching μ'. It must be that either $f_1 >_{m_1} f_a$ and $m_a >_{f_1} m_1$, or $f_a >_{m_1} f_1$ and $m_1 >_{f_1} m_a$. That is, given two individuals paired under a stable matching and the existence of an alternative stable matching, one mate prefers the matching that keeps them together, and one prefers the matching that separates them. In any stable matching where you prefer your actual mate to another attainable mate, your mate necessarily prefers some other attainable mate to you!

Strategic issues: Should I lie about my preferences?

Thus far, we have considered the conditions under which stable matchings will exist, and we have characterized some of the interesting properties of those stable matchings. We have yet to say anything about how individuals should behave strategically so as to achieve desirable matchings, stable or otherwise. In this subsection, we begin to explore questions of this sort.

Consider first the idealized class of mating systems in which all participants have complete information. That is, all participants (1) have sufficient information to rank the desirability of all members of the opposite sex, (2) know their own desirability because they know the preferences of each member of the opposite sex, and (3) know the preferences of each member of the same sex.

We begin by considering a male-courtship system, as described for the deferred acceptance algorithm. Will a male ever benefit by deceptive behaviour, displaying to a female who is not his first choice among the females to whom he has not yet displayed? Will a female ever benefit by deceptive behaviour, rejecting males who are preferred to currently engaged males, accepting suitors when the currently engaged male is preferred, remaining single despite receiving a courtship from an acceptable male, or engaging an unacceptable male despite a preference for remaining single?

Roth (1982) demonstrated that, under the male-courtship matching procedure, males will never benefit from deceptive behaviour. In game-theoretic terms, the straightforward behaviour of displaying to the first-choice female that has not yet rejected the male is a dominant strategy. Females, however, can benefit from deceptive behaviour in this system whenever the F-optimal and M-optimal matchings differ. Consider the simple female choice strategy of rejecting all but the male with whom the female would be matched under the F-optimal matching. When all females follow this strategy, under the deferred acceptance procedure with male courtship, the F-optimal matching is generated instead of the M-optimal matching. The females are in effect cooperating to force the male-courtship system to their shared interest, the F-optimal matching, and by definition some or all females are better off for having behaved 'deceptively'.

A *potentially stable matching procedure* is a procedure, such as the deferred acceptance algorithm, that will always generate a stable matching if no individual behaves deceptively. The Impossibility Theorem for two-sided matching (Bergstrom and Manning, 1982; Roth, 1982; Roth and Sotomayor, 1990) states that no potentially stable matching procedure can possibly remove all incentive for deceptive behaviour. The implication is striking: *regardless* of the mating system and matching procedure, either an unstable matching will be generated, or there will be an incentive for some individuals to behave in a circumspect manner regarding their actual mate preferences, or both. Moreover, when individuals behave deceptively, potentially stable matching procedures do not necessarily generate stable matchings. The Impossibility Theorem, therefore, limits the efficacy of matching mechanisms for generating matchings, when individuals employ strategic (sometimes deceptive) behaviours.

To this point, we have assumed that all individuals have perfect information. In practice, individuals may face imperfect information in many forms. For example, a particular female may be hindered by uncertainty about her own preferences because she is uncertain about the quality of potential partners. She may be uncertain about the distribution or identities of potential partners. She may face uncertainty about the preferences of potential partners (and thus her own 'rank', in terms of desirability). Finally, she may face uncertainty about the preferences of the other females. Without all of this information, she will be unable to compute the F-optimal matching and, therefore, be unable to employ the strategy described above that ensures the F-optimal matching under the male-courtship form of the deferred acceptance algorithm. What sort of strategic behaviour should she employ, given imperfect information in any of these forms? While this question is too involved to consider here, it has been treated in detail by Roth (1989), the results of which are summarized in Roth and Sotomayor (1990). These results may be of considerable interest to those studying mate choice tactics.

DISCUSSION

In this paper, we have outlined a theoretical framework – two-sided matching theory – that can be employed to further our understanding of mate choice systems, particularly given the prevalence of non-uniform preferences and active choice by both sexes. Within this framework, we presented an appropriate stability concept for pairing via mutual mate choice. The theory allows us to compare the relative stability of alternative mating systems, to highlight systematic conflicts of interest between the sexes, to suggest how the matching dynamics may aggravate or ameliorate these conflicts, and to generate predictions about optimal individual tactics in mutual mate choice systems.

As discussed above, the examples considered here have been concerned predominantly with systems in which all participants have perfect information and in which search is not necessary. While such an analysis provides a useful baseline for the consideration of stability and strategy in mate choice systems, in practice the participants will rarely have perfect information. Ultimately, successful application of two-sided matching theory to mate choice will require an understanding of the degree to which the results presented above hold under various forms of incomplete information.

Within two-sided matching theory, there is an inevitable conflict between the M-optimal and F-optimal matchings. In obligately sexual organisms, however, the total number of offspring produced by males is necessarily equal to the total number of offspring produced by females, in any given breeding season. Therefore, there must be an equality in fitness for males and females. While preferences are assumed to be exogenous in the economics literature, evolutionary biologists typically assume that preferences are shaped by natural selection so as to maximize the individual's reproductive fitness. If so, males and females are limited in what constitutes reasonable mating preferences. It is a major open question whether, given these constraints, there will still be an inevitable conflict of interest between the sexes, and whether there will still be incentives for deceptive individual behaviour in the matching process. Instead, natural selection may cause a convergence of the M-optimal and F-optimal strategies. We do not yet know.

As Johnstone (1997) has indicated, one of the appeals of sequential search theory is its ability to generate readily testable hypotheses that are open to empirical investigation. Often these tests can be undertaken through direct manipulation of the population (see Wiegmann *et al.*, 1996, for a thorough discussion of the testable distinguishing features of different models of sequential mate choice). In a similar vein, we hope that matching models of mate choice will generate specific testable predictions about mating markets. The simple models that we have discussed here do lead to some general hypotheses, although they are not as clearly defined as in sequential search. Nonetheless, we feel that these general hypotheses should generate some new approaches to field manipulations of mating populations. Three particular questions emerge as potential areas for empirical validation.

First, the stability of the matching can be assessed through experiments in which individuals are selectively removed from the population (e.g. Smith *et al.*, 1996; Lifjeld *et al.*, 1997; Sheldon *et al.*, 1999). If a dominant male is removed, for example, does a new matching arise where all males are simply shifted up one place in the system, or is there a complete reshuffling of positions in the hierarchy? Uniform translation upwards may indicate a system characterized by uniform preferences; a reshuffling would be more characteristic of non-uniform preferences. Blum *et al.* (1997) explore the dynamics of these *vacancy chains* in a labour market context: what happens when a senior-level employee retires and a replacement must be found internally, or lured away from another firm? The new replacement will herself vacate a position, requiring an additional search, and so forth, until the market equilibrates. Blum *et al.* briefly review prior work on the subject (sociologists have studied vacancy chains in occupations ranging from pastors and rabbis to NCAA football coaches to insane asylum superintendents) and use two-sided matching theory to model this process. Many of their results should apply equally well to vacancy chains in mating systems.

Secondly, as we have indicated, the exact stable matching generated from F-optimal choice will be different from the matching derived through M-optimal choice. Consequently, we should see major alterations in the stable matching if we could alter which sex is dominating the choice patterns. This may be possible through manipulations of one sex or the other. For example, in insect systems where males offer nuptial gifts, the male chooses females on the basis of body size and the female chooses males on the basis of the size of their nuptial gift. The dominant sex that controls choice might be manipulated by controlling the availability and distribution of nuptial gifts.

Lastly, we suggest that the overall stability of mating system pairings may be quite different depending on the type of mating system. Two-sided matching theory, for example, suggests that polygamous mating systems have a great turnover rate in pairings than serially monogamous mating systems. One avenue for testing two-sided matching theory will reside in the comparative analysis of mating system stability across different mating system types.

As a final note, we wish to echo the concluding remarks of the original paper on two-sided matching. In that paper, Gale and Shapley (1962) emphasized the *accessibility* of this type of theory. Throughout this paper, we have employed only simple mathematics – never even using calculus, let alone more sophisticated forms of analysis – and yet we have been able to describe a number of powerful and counterintuitive results from the economics literature, and to derive a few extensions of our own. Although we have not presented specific examples in detail (coalition formation, structured forms of polygamy, etc.), we stress that one can derive interesting results regarding the stability of such systems or strategic behaviour in them without employing advanced mathematical techniques. Two-sided matching theory, in essence, is based upon the careful application of structured thought to formal descriptions of matching systems.

ACKNOWLEDGEMENTS

This paper could not have been written without the help of Ted Bergstrom and Al Roth, who suggested useful inroads into the two-sided matching literature, offered numerous patient explanations and intriguing discussions, and provided valuable comments on the manuscript. Michael Lachmann and an anonymous referee also provided helpful criticism. C.B. is partially supported by NIH/NIAD grant T32-AI0742. L.R. is supported by NSF grant DEB 9629748.

REFERENCES

Alkan, A. 1986. Nonexistence of stable threesome matchings. *Math. Social Sci.*, **16**: 207–209.

Alpern, S. and Reyniers, D. 1999. Strategic mating with homotypic preferences. *J. Theor. Biol.*, **198**: 71–88.

Andersson, M.B. 1994. *Sexual Selection*. Monographs in Behavior and Ecology. Princeton, NJ: Princeton University Press.

Bakker, T.C.M., Künzler, R. and Mazzi, D. 1999. Condition-related mate choice in sticklebacks. *Nature*, **401**: 234.

Bateman, A.J. 1948. Intra-sexual selection in *Drosophila. Heredity*, **2**: 349–368.

Becker, G.S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Beletsky, L. and Orians, G. 1996. *Red-wing Blackbirds: Decision-making and Reproductive Success*. Chicago, IL: University of Chicago Press.

Bergstrom, T.C. and Bagnoli, M. 1993. Courtship as a waiting game. *J. Political Econ.*, **101**: 185–202.

Bergstrom, T.C. and Lam, D. 1989. The effects of cohort size on marriage markets in twentieth century Sweden. In *The Family, the Market, and the State in Industralized Countries* (T. Bengtsson, ed.). Oxford: Oxford University Press.

Bergstrom, T.C. and Manning, R. 1982. *Can Courtship be Cheatproof?* Ann Arbor, MI: University of Michigan Working Paper.

Blum, Y., Roth, A.E. and Rothblum, U.G. 1997. Vacancy chains and equilibration in senior-level labor markets. *J. Econ. Theory*, **76**: 362–411.

Crawford, V.P. and Knoer, E.M. 1981. Job matching with heterogeneous firms and workers. *Econometrica*, **49**: 437–450.

Cunningham, E.J.A. and Birkhead, T.R. 1988. Sex roles and sexual selection. *Anim. Behav.*, **56**: 1311–1321.

Dombrovsky, Y. and Perrin, N. 1994. On adaptive search and optimal stopping in sequential mate search. *Am. Nat.*, **144**: 355–361.

Everson, P.R. and Addicott, J.F. 1982. Mate selection strategies by male mites in the absence of intersexual selection by females: A test of six hypotheses. *Can. J. Zool.*, **60**: 2729–2736.

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

Getty, T. 1995. Search, discrimination, and selection: Mate choice by pied flycatchers. *Am. Nat.*, **145**: 146–154.

Gompper, M.E., Gittleman, J.L. and Wayne, R.K. 1997. Genetic relatedness, coalitions and social behaviour of white-nosed coatis, *Nasua narica. Anim. Behav.*, **53**: 781–797.

Janetos, A. 1980. Strategies of female choice: A theoretical analysis. *Behav. Ecol. Sociobiol.*, **7**: 107–112.

Jennions, M.D. and Petrie, M. 1997. Variation in mate choice and mating preferences: A review of the causes and consequences. *Biol. Rev.*, **72**: 283–327.

Johnstone, R.A. 1997. The tactics of mutual mate choice and competitive search. *Behav. Ecol. Sociobiol.*, **40**: 51–59.

Johnstone, R.A., Reynolds, J.D. and Deutsch, J.C. 1996. Mutual mate choice and sex differences in choosiness. *Evolution*, **50**: 1382–1391.

Jones, I.L. and Hunter, F.M. 1993. Mutual sexual selection in a monogamous seabird. *Nature*, **362**: 238–239.

Jouventin, P., Lequette, B. and Dobson, F.S. 1999. Age-related mate choice in the wandering albatross. *Anim. Behav.*, **57**: 1099–1106.

Kelso, A.S.J. and Crawford, V.P. 1982. Job matching, coalition formation, and gross substitutes. *Econometrica*, **50**: 1483–1504.

Knuth, D.E. 1976. *Stable Marriage and its Relation to other Combinatorical Problems*. CRM Proceedings and Lecture Notes Vol. 10, English language edition. Providence, RI: American Mathematical Society. Originally published in French under the title *Mariages stables et leurs relations avec d'autres problèms combinatoires*.

Kraak, S.B.M. and Bakker, T.C.M. 1998. Mutual mate choice in sticklebacks: Attractive males choose big females, which lay big eggs. *Anim. Behav.*, **56**: 859–866.

Lifjeld, J.T., Slagsvold, T. and Ellegren, H. 1997. Experimentally induced sperm competition in pied flycatchers: Male copulatory access and fertilization success. *Anim. Behav.*, **53**: 1225–1232.

Manning, J.T. 1980. Sex ratio and optimal male time investment strategies in *Asellus aquaticus* L. and *A. meridianus* Racovitsza (Crustacea: Isopoda). *Behaviour*, **74**: 265–273.

McNamara, J.M. and Collins, E.J. 1990. The job search problem as an employer–candidate game. *J. Appl. Probability*, **28**: 815–827.

McVitie, D.G. and Wilson, L.B. 1970. Stable marriage assignments for unequal sets. *BIT*, **10**: 295–309.

Mongell, S. and Roth, A.E. 1991. Sorority rush as a two-sided matching mechanism. *Am. Econ. Rev.*, **81**: 441–464.

Packer, C., Gilbert, D.A., Pusey, A.E. and O'Brien, S.J. 1991. A molecular genetic analysis of kinship and cooperation in African lions. *Nature*, **351**: 562–565.

Parker, G.A. 1983. Mate quality and mating decisions. In *Mate Choice* (P. Bateson, ed.), pp. 141–166. Cambridge: Cambridge University Press.

Penn, D.J. and Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.*, **153**: 145–164.

Real, L.A. 1990. Search theory and mate choice: I. Models of single-sex discrimination. *Am. Nat.*, **136**: 376–404.

Real, L.A. 1991. Search theory and mate choice: II. Mutual interaction, assortative mating, and equilibrium variation in male and female fitness. *Am. Nat.*, **138**: 901–917.

Roth, A.E. 1982. The economics of matching: Stability and incentives. *Math. Operations Res.*, **7**: 617–628.

Roth, A.E. 1984. The evolution of the labor market for medical interns and residents: A case study in game theory. *J. Political Econ.*, **92**: 991–1016.

Roth, A.E. 1989. Two-sided matching with incomplete information about others' preferences. *Games Econ. Behav.*, **1**: 191–209.

Roth, A.E. 1990. New physicians: A natural experiment in market organization. *Science*, **250**: 1524–1528.

Roth, A.E. 1991. A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *Am. Econ. Rev.*, **81**: 415–440.

Roth, A.E. and Sotomayor, M.A.O. 1990. *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis*. Econometric Society Monographs No. 18. New York: Cambridge University Press.

Roth, A.E. and Vande Vate, J.H. 1990. Random paths to stability in two-sided matching. *Econometrica*, **58**: 1475–1480.

Shapley, L.S. and Shubik, M. 1972. The assignment game I: The core. *Int. J. Game Theory*, **1**: 111–130.

Sheldon, B.C., Davidson, P. and Lindgren, G. 1999. Mate replacement in experimentally widowed collared flycatchers (*Ficedula albicollis*): Determinants and outcomes. *Behav. Ecol. Sociobiol.*, **46**: 141–148.

Smith, H.G., Wennerberg, L. and von Schantz, T. 1996. Sperm competition in the European starling (*Sturnus vulgaris*): An experimental study of mate switching. *Proc. Roy. Soc. Lond. B*, **263**: 797–801.

Sterck, E.H.M., Watts, D.P. and van Schaik, C.P. 1997. The evolution of female social relationships in nonhuman primates. *Behav. Ecol. Sociobiol.*, **41**: 291–309.

Tamura, A. 1993. Transformation from arbitrary matchings to stable matchings. *J. Combinatorial Theory*, *Ser. A*, **62**: 310–323.

Waterman, J.M. 1997. Why do male Cape ground squirrels live in groups? *Anim. Behav.*, **53**: 809–817.

Watts, D.P. 1998. Coalitionary mate guarding by male chimpanzees at Ngogo, Kibale National Park, Uganda. *Behav. Ecol. Sociobiol.*, **41**: 43–55.

Widemo, F, and Sæther, S.A. 1999. Beauty is in the eye of the beholder: Causes and consequences of variation in mating preference. *Trends Ecol. Evol.*, **14**: 26–31.

Wiegmann, D.D., Real, L.A., Capone, T.A. and Ellner, S. 1996. Some distinguishing features of models of search behavior and mate choice. *Am. Nat.*, **147**: 188–204.

APPENDIX

Polygamous mating systems need not allow any stable matchings, when the number of females paired to each male can vary across matchings. Roth and Sotomayor (1990) give the following illustration (example 2.7 in their book). Consider a system with two males $M = \{m_1, m_2\}$ and three females $F = \{f_1, f_2, f_3\}$. Females are concerned *only* with the identity of the male to whom they are matched (and not with how many other mates that male has). Females f_1 and f_2 prefer m_2 to m_1, while female f_3 prefers m_1 to m_2. Males desire at most two mates, and have preferences $\{f_1, f_3\} >_{m_1} \{f_1, f_2\} >_{m_1} \{f_2, f_3\} >_{m_1} \{f_1\} >_{m_1} \{f_2\}$ and $\{f_1, f_3\} >_{m_2} \{f_2, f_3\} >_{m_2} \{f_1, f_2\} >_{m_2} \{f_3\} >_{m_2} \{f_1\} >_{m_2} \{f_2\}$. The reader can verify that no stable matching exists in this system (see Roth and Sotomayor, 1990, for details).

Even when the number of females paired to each male is fixed, stable matchings need not exist, if females are concerned with the identities of the other mates of the male to whom they are paired. For example, consider a system with two males $M = \{m_1, m_2\}$ and four females $F = \{f_1, f_2, f_3, f_4\}$, where, by the matching rules, each male takes exactly two mates. The two males have identical preferences, preferring females in the order $f_1 > f_2 > f_3 > f_4$. Females prefer any pairing to remaining single, but are more concerned with the identify of the other female with whom they share a male than with the identify of the male himself: f_1 prefers to share either male with f_2 to any other arrangement. Similarly, f_2 prefers to be with f_3, f_3 with f_4, and f_4 with f_1. Although these conditions fully specify neither male nor female preferences, they are sufficient to guarantee that a blocking pair always exists. Note that any male not paired to both f_1 and f_2 will willingly form a blocking pair with either of these females. When f_1 does not share a male with f_2, she will abandon her mate for the other male, to be with f_2. When f_1 does share a male with f_2, then f_2 will abandon this mate for the other male, to be with f_3.