REVISITING THE VARIANCE-BASED SELECTION MODEL OF DIPLOID DRONE PRODUCTION FOR MULTIPLE MATING IN HONEY BEES

M. CHHETRI, N. JOHNSON, O. RUEPPELL, AND J. RYCHTÁŘ

Abstract. The high number of mates of honeybee queens has lead to the proposal of several adaptive explanations. The competing hypotheses to explain multiple mating in honeybees and some other social insects have been mostly evaluated empirically with comprehensive theoretical analysis lacking behind. We report on the mathematical analysis of the diploid drone hypothesis for multiple mating, which suggests that multiple mating evolved as a safeguard against the production of infertile male offspring. In accordance with earlier models, our analysis shows that multiple mating does not reduce the average value of diploid drone production but reduces its variance. We combine this observation with a colony growth model to assess the impact of this reduction in variance to the colony fitness. Considering a plausible parameter space for the honeybee, we conclude that the reduction in variance of diploid drone production can be a significant selective force for multiple mating.

We have also described rules of a game for which a problem of finding the best strategy is equivalent to the above biological problem of bee mating. We made a significant progress in the general solution of this game and conjectured that the best strategy is strongly related to the geometry of rational numbers.

1. Introduction

There are several competing hypothesis explaining the multiple mating of western honeybees (*Apis mellifera*). In this paper we focus on a diploid drone production hypothesis. It predicts that multiple mating reduces a production of diploid drones - sterile males whose production means costs to the bee colony without providing any benefits in returns. There has been no comprehensive theoretical analysis of this hypothesis so far.

We build an explicit ODE model for colony growth (Section 3), taking the diploid drone production as well as other biological phenomena (summarized in Section 2) into account. Using this ODE model and binomial distribution, we introduce two fitness functions (Section 4) that measure the colony success as a function of a number of mates. This yields an optimization problem - find the number of mates for which the fitness is maximal. In Section 5 we reintroduce the problem with the help of an abstract game. Finding the best strategy for our game is equivalent to the above biological problem of bee mating. We made a significant progress in the general solution of this game and conjectured that the best strategy is strongly related to the geometry of rational numbers.

Date: January 2, 2008.

Key words and phrases. ODE model, Binomial distribution, variance, optimization, guessing game, geometry of rational numbers.

The research was supported by the NSF grant 0634182.

For a large number of mates, the fitness functions are maximized qualitatively in Section 6. Also, in Section 9, we approximate the binomial distribution by normal distribution. This approximation made theoretical analysis more feasible and we were able to confirm results obtained in Section 6.

For a small number of mates, the fitness functions are maximized quantitatively in Section 7 . We have discovered that the behavior of the fitness functions depends heavily on the properties of rational numbers. In Section 8, we discuss in detail the impact of the geometry of rational numbers on our analysis and present a conjecture that might be of interest to both mathematicians and biologists.

2. Biological background

Mating patterns across the animal kingdom are highly variable but most species are either polygamous (multiple males mate with multiple females) or polygynous (a single male mating with several females). Monogamy and polyandry (one female mating with multiple males) are much rarer (Andersson 1994). Within the social insects monogamy is the rule but numerous exceptions for polyandry exist. Queens of honeybees (Palmer and Oldroyd 2000), leaf-cutter ants (Boomsma et al. 1999), harvester ants (Wiernasz et al. 2004), old- and new-world army ants (Denny et al. 2004; Kronauer et al. 2006), and higher vespid wasps (Foster and Ratnieks 2001) mate with multiple males, often to extreme degrees (Wattanachaiyingcharoen et al. 2003). These mating occur during mating flights at the beginning of their reproductive lives on one or several mating flights, and the sperm is stored for the rest of their lives to reproduce new worker and sexuals. Re-mating seems not to occur (Kronauer et al. 2006).

In western honeybees (*Apis mellifera*), the reproductive cycle starts with the raising of a set of new queens from existing brood in the hive. Before these new queens emerge and fight each other until only one remains (Tarpy et al. 2004), the old queen leaves the hive with about half of the workers to establish a new colony elsewhere (reproductive swarming). The single remaining new, virgin queen engages in one to five mating flights (Schluns et al. 2005) during which she mates with up to 45 males (Neumann et al. 1999). After her final mating flight the newly mated queen starts to produce new offspring. Normally, the majority of the offspring consists of new workers to increase the workforce of the hive to ensure successful survival (Winston 1987).

Production of workers is under queen control because honeybees exhibit a haplodiploid sex determination system: Unfertilized (haploid) eggs develop into males, fertilized (diploid) eggs into females that mostly develop into workers. The molecular basis for this sex determination has been identified as the csd gene that governs the sex determination cascade in honey bees (Beye et al. 2003). If this gene is present in two different versions (alleles), the organism develops into a female, if it is represented only by one allele, male development results. Consequently, males develop from haploid eggs but also from diploid eggs that bear accidentally two copies of the same allele (Beye et al. 2003). These diploid males are infertile and generally cannibalized within 72 hours of emergence (Woyke 1963). Diversifying selection has generated allelic diversity at the *csd* locus with allele number estimates varying between 11 and 19 (Hasselmann and Beye 2004). However, a queen mating with only one male with an identical csd allele (incompatible drone) will suffer 50% mortality of its diploid (worker destined) offspring. In contrast, when only one of two males bears an identical *csd* allele, diploid drones will only make up 25% of all workers destined brood. Upon survival to the new reproductive season, the focal hive will produce usually one swarm and numerous males that attempt to mate with virgin queens from other hives. In some cases, large original swarms can produce secondary after-swarms but the fitness impact of this is debatable (Winston 1987).

Presumably, multiple mating bears general costs for females, including time and energy expenditure, and increased risks of predation and infection with sexually transmitted diseases (Brown and Schmid-Hempel 2003). Potential benefits are equally numerous and several hypotheses have been suggested to explain female multiple mating in the social Hymenoptera (Palmer and Oldroyd 2000; Crozier and Fjerdingstad 2001). Mating for sufficient sperm has been proposed (Cole 1983) but largely dismissed for honeybees (but see (Schluns et al. 2005)). More emphasis has recently been placed on the fact that multiple mating increases the genetic heterogeneity within the hive. Increased genetic variability may have positive consequences for the disease resistance within hives (Brown and Schmid-Hempel 2003) and enhance the division of labor and thus maintain nest homeostasis (Oldroyd and Fewell 2007). Both suggestions have been supported by empirical evidence (Oldroyd and Fewell 2007; Seeley and Tarpy 2007). In addition, multiple mating may reduce the risk of producing a large proportion of inviable brood to diploid male production (Page 1980). After the identification of the sex determination mechanism (Beye et al. 2003; Hasselmann and Beye 2004), this suggestion could be precisely tested on a large scale and we therefore extend the initial analysis of Page (1980) with a new mathematical model that connects a colony growth model of ordinary differential equations with a fitness step function (Sherman et al. 1988; Tarpy and Page 2002), based on the observation that a colony either reaches a critical size to reproduce or fails after establishment. We subsequently link mating frequency, the number of alleles in the population, and the critical colony size for success to queen fitness.

The previous model of variance-based selection for honeybee multiple mating due to diploid drone production (Page 1980) built on empirical observations of the quasi-logistic growth of honey bee colonies (Sakagami and Fukuda 1968) to model colony growth with a logistic function. It uses a second logistic function to link the fitness of the queen to the proportion of viable brood produced which in turn depends on the number and compatibility of the queen's mates (Page 1980). The combination of these two model components lead to the conclusion that multiple mating continuously increases individual queen fitness in populations with multiple alleles at the csd locus and benign conditions (i.e. a relatively high brood mortality can be tolerated) (Page 1980).

In general, our results are in accordance with those reported by Page (1980). However, our analysis emphasize that honeybee multiple mating has quickly diminishing returns, and that fitness functions are not continuously increasing with increased number of mates. Instead, discontinuous step functions produce results that argue against selection for multiple mating by diploid drone production in a significant portion of the parameter space. This effect depends on the continuously increasing risk of mating with at least n incompatible drones, relative to the stepwise increase in the probability of growing to the critical colony size.

We believe that the detailed predictions of our analysis will generate new empirical investigations into the importance of diploid drone production, which may be of considerable practical and theoretical importance, given the highly structured honeybee populations.

3. Colony Growth Model

For our model we only count workers towards the colony size, C. This can be justified since drones only consume resources before they mate and die, or are "kicked out" of the hive at the end of the mating season (Fukuda and Ohtani, 1977). We use the general growth model:

$$C'(t) = (b-d)C$$

where $b > 0$ is the birthrate and $d > 0$ the death rate of workers.

Lifespan of honey bees is different between fall and spring season, and summer season. On average, it is 30-60 days for fall and spring and 15-38 days for summer (Winston, 1987). We set the death rate, d, to be the inverse of the lifespan.

We assume honey bee queens lay a number of eggs proportional to the size of the colony, C, divided by $h+k$, where h is the number of workers needed to forage for, and k is the number of workers needed to take care of, each uncapped larva properly. Both h and k are 5 on average for European honey bees (Eishchen et al., 1982). Workers take care of the larva for a time T (5.5 days on average in European honey bees (Tribe & Fletcher, 1977)). Also, we put an upper bound on the queen egg-laying rate, B (2000 eggs per day in European honeybees (Camazine, 1991)).

This gives us the following:

$$C'(t) = \begin{cases} \left(\frac{1}{(k+h)T} - d\right) C, & \frac{C}{(k+h)T} < B \\ B - dC, & \text{o/w.} \end{cases}$$

Not all offspring destined workers actually become workers due to the possibility that some eggs develop into diploid drones (half of the offspring develop into diploid drones sons if a father drone carries the same allele at the csd locus as the queen). If the queen mates with n drones, from that mated with R_n incompatible drones, then the laying rate will be reduced by a factor $1 - \frac{R_n}{2n}$ since out of the all possible eggs, $\frac{R_n}{2n}$ will give rise to diploid drones instead of workers.

Consequently, the colony growth is given by

$$C'(t) = \begin{cases} \left(\frac{1}{(k+h)T} - d\right) C, & \frac{C}{(k+h)T} < B\left(1 - \frac{R_n}{2n}\right) \\ B\left(1 - \frac{R_n}{2n}\right) - dC, & \text{o/w.} \end{cases} \tag{3.1}$$

The non-equilibrium solutions of (3.1) tend to equilibriums exponentially fast and the only stable equilibrium of the system is

$$C_c = \frac{B\left(1 - \frac{R_n}{2n}\right)}{d} \tag{3.2}$$

which is also called the carrying capacity.

4. Colony fitness

The fitness is the capability of an individual (in our case the whole colony) of a certain genotype to reproduce, and it is defined as the proportion of the individual's genes in all the genes of the next generation. We are here interested in the relative fitness which is quantified as the average number of surviving progeny of a particular genotype compared with average number of surviving progeny of competing genotypes after a single generation, i.e. one genotype is normalized at fitness 1 and the fitnesses of other genotypes are measured with respect to that genotype. Relative fitness can therefore take any nonnegative value, including 0.

In order to swarm, the colony needs to grow to a certain critical size C_c (Winston, 1987). We measure the colony fitness as a probability that it will reach such a size. That is, we measure the relative fitness of the colony comparing it to an ideal colony that reaches the critical size C_c every time.

4.1. **Preliminary calculation of the colony fitness function.** Firstly note that, by (3.2), the carrying capacity is dependent on the number of mates only through the term $\left(1-\frac{R_n}{2n}\right)$, the proportion of genetically healthy workers in the offspring, and that the dependence on this factor is linear. We thus get

$$\begin{aligned} P(C \geq C_c) &= P\left(C \geq \frac{B\left(1-\frac{R_n}{2n}\right)}{d}\right) \\ &= P\left(\frac{R_n}{n} \leq 2-\frac{2C_c d}{B}\right). \end{aligned}$$

Letting

$$\Omega_n = \frac{R_n}{n}$$

be the real fraction of incompatible drones the queen mated with, we can interpret

$$\Theta = 2\left(1-\frac{C_c d}{B}\right)$$

as the tolerance level. Thus we measure the colony fitness as a probability that the queen did not encounter more incompatible drones than what she is able to tolerate.

4.2. **Binomial distribution for R_n.** Queen bees mate with drones in the drone congregation areas (areas with very high drone density: Baudry et al. 1998). After mating, queens store drones' semen in an organ called the spermatheca. A drone always dies after mating once. After a queen has mated with as many drones as she wanted (or could), she returns to the hive and begins to lay eggs. A full spermatheca holds 5.4 to 5.7 million sperm which is much more than the number of eggs the queen lays (Winston 1987). We assume that each drone contributes an equal amount of sperm.

We need to calculate Ω_n, the proportion of incompatible drones the queen mated with if she mated with n drones in total. Definitely, Ω_n depends on the real proportion of incompatible drones in the population, which we denote by Ψ. If A is the number of alleles in the csd locus (A is estimated between 11 and 19, Hasselmann and Beye 2004), then

$$\Psi = \frac{2}{A}. \tag{4.3}$$

Queen bears 2 different alleles, drone bears only one. For incompatible individuals, the likelihood is high that the drone's allele matches one of the queen's alleles.

Given the high number of drones in the drone congregation area and the randomness of their origin, we may assume that the probability the queen mated with r incompatible drones, provided it mated with n drones in total, is given by the binomial distribution with parameters n and Ψ, i.e. by

$$\binom{n}{r}\Psi^r(1-\Psi)^{n-r}. \tag{4.4}$$

This in turn means that Ω_n is a random variable attaining values $\frac{r}{n}$; $0 \le r \le n$, with probability $\binom{n}{r}\Psi^r(1-\Psi)^{n-r}$.

4.3. **Formulas for Colony Fitness.** Putting the preliminary calculations together yields the fitness as a function of n, Ψ and Θ, specifically,

$$\begin{aligned} F(n,\Psi,\Theta) &= P(\Omega_n \le \Theta) = P(R_n \le n\Theta) \\ &= \sum_{r=0}^{\lfloor n\Theta \rfloor} \binom{n}{r}\Psi^r(1-\Psi)^{n-r} \end{aligned} \tag{4.5}$$

We also consider the function

$$G(n,\Psi,\Theta) = \begin{cases} F(1,\Psi,\Theta), & n = 1, \\ \frac{1}{3}\sum_{i=-1}^{1} F(n+i,\Psi,\Theta), & n > 1. \end{cases} \tag{4.6}$$

The equation (4.6) is a special case of a more general form

$$G(n,\Psi,\Theta) = \sum_{i=1}^{\infty} p_i^{(n)} F(i,\Psi,\Theta) \tag{4.7}$$

where $p_i^{(n)} \in [0,1]$ and

$$\sum_{i=1}^{\infty} i p_i^{(n)} = n.$$

The introduction of this function G is justified by the fact that often the queen (aims) for n mates but practically can achieve any number of mates. If she aims for n, she achieves i with probability $p_i^{(n)}$ and achieves n “on average”. This is a biologically reasonable model since honey bee queen presumably do not count the number of mates exactly but likely can implement some mechanism to be able to “count” approximately (Tarpy and Page 2000).

We study in details only the function G given by (4.6). The more general fitness given by (4.7) yields comparable results as long as the distribution given by $p_i^{(n)}$ is biologically meaningful, i.e. the probable number of mates are somehow centered around the targeted number of mates.

5. MATH ONLY FORMULATION - GUESSING GAME

Recall that our main goal is to maximize the fitness function

$$F(n,\Psi,\Theta) = \sum_{r=0}^{\lfloor n\Theta \rfloor} \binom{n}{r}\Psi^r(1-\Psi)^{n-r},$$

that is, for any given Ψ and Θ, find n such that the value of $F(n,\Psi,\Theta)$ is maximal. Additionally, we are interested in finding n (possibly different from above) such that the value of $G(n,\Psi,\Theta)$, given by (4.6), is maximal.

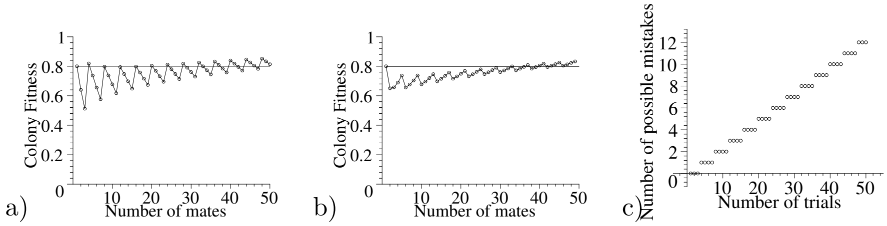

FIGURE 1. a) $F(n, 0.2, 0.25)$, b) $G(n, 0.2, 0.25)$, c) $\lfloor n \cdot 0.25 \rfloor$

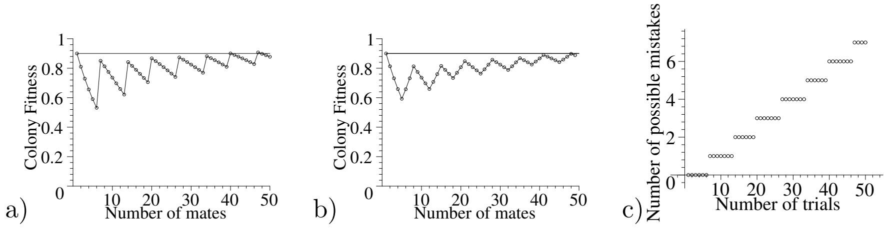

FIGURE 2. a) $F(n, 0.1, 0.15)$, b) $G(n, 0.1, 0.15)$, c) $\lfloor n \cdot 0.15 \rfloor$

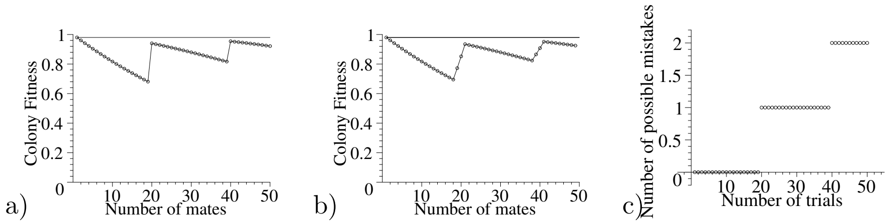

FIGURE 3. a) $F(n, 0.02, 0.05)$, b) $G(n, 0.02, 0.05)$, c) $\lfloor n \cdot 0.05 \rfloor$

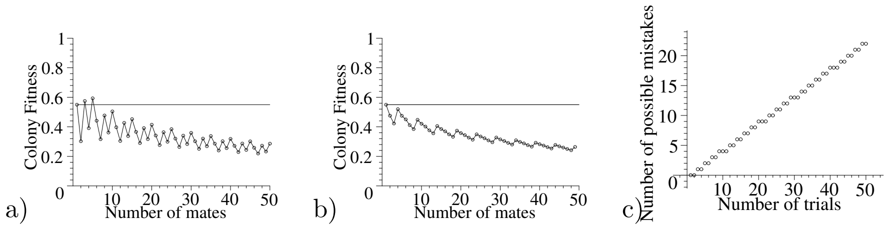

FIGURE 4. a) $F(n, 0.45, 0.4)$, b) $G(n, 0.45, 0.4)$, c) $\lfloor n \cdot 0.4 \rfloor$

In this subsection, we reintroduce the problem and state it in a mathematical setting. For that, we need to consider the following game.

GUESSING GAME *The game is played in rounds. During each round, a player has to answer a yes/no question and then decide whether to continue or to quit the game. If he continues, he plays another round of the game. If he quits (or continues*

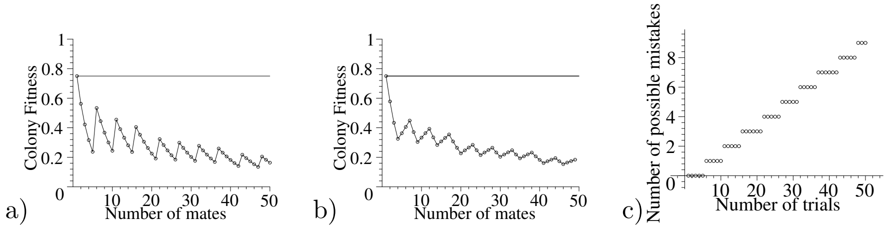

FIGURE 5. a) $F(n, 0.25, 0.19)$, b) $G(n, 0.25, 0.19)$, c) $\lfloor n \cdot 0.19 \rfloor$

indefinitely), he receives a payoff of 1 *provided he answered less than* $\Theta \cdot 100\%$ *of questions incorrectly and receives a payoff of* 0 *otherwise.*

Provided a player does not know the answers and can only guess (with the probability of making wrong guess Ψ), the player's strategy consists of the choice of n, the number of rounds he will play the game. For a fixed n, the expected player's payoff $F(n, \Psi, \Theta)$ is given exactly by (4.5). The question is, what strategy should the choose to maximize the payoff.

We also introduce the following variant of the GUESSING GAME, call it BOUNDED GUESSING GAME, with the additional restriction that the player has to stop no later than at a given round N (determined at the beginning of the game). If we wish to specify the number N, we refer to it as N-BOUNDED GUESSING GAME.

6. QUALITATIVE ANALYSIS - BEHAVIOR FOR LARGE n

For large n, we can derive relevant properties and behaviors of the functions $F(n, \Psi, \Theta)$ and $G(n, \Psi, \Theta)$ without using the explicit formulas (4.5) and (4.6). Indeed, since R_n is binomially distributed and $\Omega_n = \frac{R_n}{n}$, it follows from the standard properties of binomial distribution that

$$E(\Omega_n) = E\left(\frac{R_n}{n}\right) = \frac{E(R_n)}{n} = \Psi \tag{6.8}$$

and

$$\operatorname{var}(\Omega_n) = \operatorname{var}\left(\frac{R_n}{n}\right) = \frac{\operatorname{var}(R_n)}{n^2} = \frac{\Psi(1-\Psi)}{n}. \tag{6.9}$$

Moreover, using Hölder inequality, (6.8) and (6.9), we have

$$\begin{aligned} E\big|\Omega_n - \Psi\big| &= E\Big|\Omega_n - E(\Omega_n)\Big| \le \sqrt{E\left[\Big(\Omega_n - E(\Omega_n)\Big)^2\right]} \\ &= \sqrt{\operatorname{var}(\Omega_n)} = \sqrt{\frac{\Psi(1-\Psi)}{n}}. \end{aligned}$$

Hence

$$\lim_{n\to\infty} E\big|\Omega_n - \Psi\big| = 0. \tag{6.10}$$

Consequently, since

$$E\big|\Omega_n - \Psi\big| \ge \varepsilon P\Big(\big|\Omega_n - \Psi\big| \ge \varepsilon\Big)$$

we have, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\Big(|\Omega_n - \Psi| \geq \varepsilon\Big) = 0. \tag{6.11}$$

The equation (6.11) proves the following:

Theorem 1. *For any* $\varepsilon > 0$ *and any threshold probability* $\pi_0 < 1$*, there exists* n_0 *such that* $|\Omega_n - \Psi| < \varepsilon$ *with probability at least* π_0*, for all* $n > n_0$*.*

Corollary 2.

$$\lim_{n\to\infty} F(n,\Psi,\Theta) = \begin{cases} 1, & \textit{if } \Psi > \Theta, \\ 0, & \textit{if } \Psi < \Theta, \\ 0.5, & \textit{if } \Psi = \Theta. \end{cases}$$

Same is true for $G(n,\Psi,\Theta)$*.*

Proof. The first two cases follow directly from the Theorem 1. The case when $\Psi = \Theta$, is an easy consequence of the fact that R_n has a binomial distribution and thus its median is one of the following three numbers $\{\lfloor np\rfloor - 1, \lfloor np\rfloor, \lfloor np\rfloor + 1\}$. □

7. QUANTITATIVE ANALYSIS - BEHAVIOR FOR SMALL n

The conclusion of Corollary 2 for a queen bee is that if $\Psi < \Theta$, i.e. when there is less incompatible drones than she can tolerate, she should mate with a very high number of drones. However, neither this result quantifies this "high number" nor it provides with any idea as to what the queen should do if $\Psi \leq \Theta$. It only says that if $\Psi \leq \Theta$, then mating with a large number of drones may decrease the fitness to almost 0.

From the proof of the Theorem 1, one could get a quantitative estimate on n_0, the number of mates that would guarantee the probability of $|\Omega_n - \Psi| < \varepsilon$ being as close to 1 as needed. Yet, n_0 would be so large that it would be practically impossible for a queen to mate with so many drones. In this section, we try to investigate some properties of $F(n,\Psi,\Theta)$ in order for us to predict what a queen should do to maximize the fitness while mating with only finite (relatively small and realistic) number of mates.

7.1. **Local maxima of** $F(n,\Psi,\Theta)$**.** It can be seen from figures (1(a)-5(a)) that the fitness function F is very jumpy. The function is never increasing or decreasing on a sufficiently long interval. Yet one can speak about increasing and decreasing tendencies from the same set of figures. The same is true about function G, figures (1(b)-5(b)), although the function G is "smoother" than F (one can see from figures that G has smaller oscillation than F). Observe that both functions have quickly diminishing returns.

The local maxima of the function F (unlike G) can be determined analytically. In fact, we have the following theorem:

Theorem 3. *For given* $\Psi \in (0,1)$ *and* $\Theta \in (0,1)$*, the function* $F(n,\Psi,\Theta)$ *has a local maxima at* $n = k$ *if and only if*

$$\lfloor (k-1)\Theta\rfloor < \lfloor k\Theta\rfloor = \lfloor (k+1)\Theta\rfloor. \tag{7.12}$$

Proof. The statement is evident when considering the guessing game. In order to get a nonzero payoff, the player playing k rounds can make as much as $m = \lfloor k\Theta\rfloor$

Range for Θ	Range for Ψ	$\text{argmax}_{n\le 20}F(n,\Psi,\Theta)$
$(0,1/15)$	$(0,1/2]$	1
$(1/15,1/3)$	$(\Theta-0.05,1/2]$	1
$(1/15,1/3)$	$(0,\Theta-0.05)$	$q\in q(\Theta)$
$(1/3,1/2]$	$(0,1/2)$	$q\in q(\Theta)$
$(1/3,1/2)$	$1/2$	1
$1/2$	$1/2$	2

Table 1. Location of absolute maxima of $F(n,\Psi,\Theta)$ on $n\le 20$; $q(\Theta)$ is a set of integers $q\le N=20$ for which there is an integer p such that $0\le\Theta-p/q$ is small.

mistakes. If k satisfies the conditions (7.12), the player can make $m-1$ mistakes in $k-1$ rounds, and m mistakes in $k+1$ rounds.

Clearly, $F(k,\Psi,\Theta)>F(k+1,\Psi,\Theta)$ because whenever one makes only m mistakes in $k+1$ rounds, then he made at most m mistakes in k rounds. The inequality is strict, because there is a nonzero chance of making a mistake in the $k+1$ round and thus making actually only $m-1$ mistakes in the first k rounds.

Similarly, $F(k-1,\Psi,\Theta)<F(k,\Psi,\Theta)$. Indeed, if one manages to make no more than $m-1$ mistakes in $k-1$ rounds, one can make a mistake in the kth round and still not make more than m mistakes in k rounds. Since the probability of making a mistake in the kth round is less than 1, the inequality is strict. □

7.2. **Global behavior.** To find an absolute maxima of $F(n,\Psi,\Theta)$; $n\in\{1,\cdots,N\}$ for general N,Ψ and Θ is analytically very difficult and we are only able to find results using computer calculations.

The major difficulty is that even if $\Psi<\Theta$, so that $F(n,\Psi,\Theta)\to 1$ as $n\to\infty$, F may initially have a decreasing tendency (see figures 2(a)-3(a)). Likewise, even if $\Psi>\Theta$, and thus $F(n,\Psi,\Theta)\to 0$ as $n\to\infty$, F may initially have an increasing tendency (see figure 4(a)).

We used Maple to calculate $F(n,\Psi,\Theta)$ and $G(n,\Psi,\Theta)$ for $n\in\{1,\dots,20\}$, and $\Psi,\Theta\in\{i/100; i=1,\dots,50\}$. From those data we calculated

$$\text{argmax}_{n\le 20}F(n,\Psi,\Theta),\ \text{and}\ \text{argmax}_{n\le 20}G(n,\Psi,\Theta)$$

for all of the above Ψ and Θ. Our results are summarized in the tables below.

The interesting feature (and a complication) is that the value of Θ does not determine the integer in the set

$$q_\varepsilon(\Theta)=\Big\{q\in\{1,\dots,N\}; \exists p \text{ such that } 0\le\Theta-p/q<\varepsilon\Big\}$$

uniquely. For example, $1/3=2/6=3/9=\cdots$, i.e. $q_\varepsilon(1/3)$ contains any multiple of 3 smaller than 20. Moreover, there are many rational numbers close to any given number. For example, 0.27 is just above $1/4, 4/15$ and $5/19$, with each of the fractions having denominator below 20. Thus, the table 1apparently does not determine the maxima. The appropriate choice of ε and $q\in q_\varepsilon(\Theta)$ depends on $\Theta-\Psi$. The larger the quantity $\Theta-\Psi$, the larger q should be. Moreover, ε should be large enough to ensure $q_\varepsilon(\Theta)\neq$ and yet not too large to have too many elements in $q_\varepsilon(\Theta)$. Some results and conjectures regarding the choice of ε are presented in the subsection 8.

Range for Θ	Range for Ψ	$\text{argmax}_{n\leq 20} G(n,\Psi,\Theta)$
$(0, 2/5]$	$(\Theta - 0.05, 1/2]$	1
$(0, 2/5]$	$(0, \Theta - 0.05]$	20
$(2/5, 1/2]$	$(\Theta, 1/2]$	1
$(2/5, 1/2]$	$(0, \Theta]$	20

TABLE 2. Location of absolute maxima of $G(n, \Psi, \Theta)$.

Notice the surprising result that once $\Theta \geq 1/3$, Ψ can be even larger than Θ and still the maximum is attained at $n > 1$.

Since G does not oscillate as much as F, the results for G does not depend so heavily on the actual value of Θ, rather they depend more on the size of $\Theta - \Psi$. The results are summarized in the table 2.

We assumed that the queen cannot mate with more than 20 mates, so 20 actually means - mate with as many as possible. There are some minor exceptions to the above table 2.When $\Theta \approx 0.25$ or $\Theta \approx 0.4$, the function F oscillates so much that even the function G oscillates quite a lot and thus the results for F partially holds for G as well.

Nevertheless, notice that if $\Theta < \Psi$, then the queen should mate with only 1 drone, while if $\Theta > \Psi$, the queen should mate with 20. Thus, these results are in agreement with the Corollary 2.

8. GUESSING GAME AND THE GEOMETRY OF RATIONAL NUMBERS

As seen in subsection 7.2, the behavior of $F(n, \Psi, \Theta)$ for relatively small n is governed mostly by Θ. We fix N and consider N-bounded guessing game (or equivalently, assume that the queen cannot mate with more than N mates). We want to find an absolute maxima of F for $n \leq N$.

The first thing we observe is that once N is fixed, the proportion of mistakes, Ω_n, is in a finite (and thus a discrete) set

$$S = \left\{ \frac{p}{q}, p \leq q \leq N \right\}.$$

Moreover, the numbers, p/q, that are below Θ but not too far below from Θ are of significance since these are the real tolerance level a player can count on. For example, if the tolerance level is $\Theta = 0.27$ and $N \geq 20$, then one can tolerate up to 5 out of 19 mistakes ($5/19 \approx 0.263$); if $N < 19$, one can tolerate only up to 1 out of 4 mistakes ($1/4 = 0.25$). In an extreme case, the decrease in the tolerance can be as high as $1/N$. Clearly, smaller the N, higher the decrease in the real tolerance level.

We can see the example of such big decrease when $\Theta < 1/N$ and yet $\Theta \approx 1/N$. In this case, one cannot make a single mistake in all of the N possible trials, i.e. the real tolerance level is 0.

The decrease of the tolerance not only depends on N, it depends on Θ as well. If the rational numbers from S below Θ are porous (such as we seen is the case of Θ just below $1/N$), the decrease in the tolerance is substantial.

The figures 6 depicts the set S of rational numbers with denominators less then N, for $N = 20, 50$ and 100. We can see that the regions with low density (white

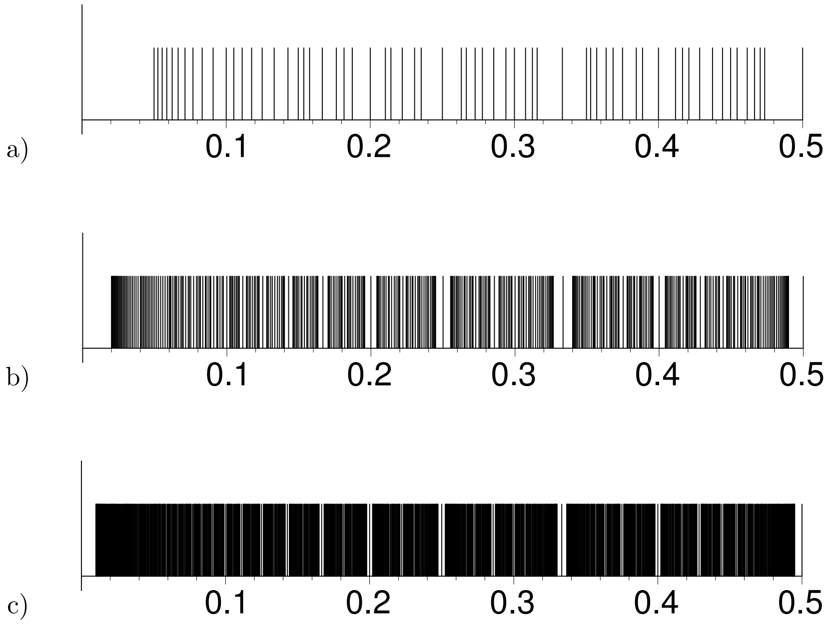

FIGURE 6. Rational numbers p/q for $q \leq N$ in $[0, 0.5]$ (in $[0.5, 1]$ is the situation symmetrical). a) $N \leq 20$, b) $N \leq 50$, c) $N \leq 100$.

spots) are exactly the tolerance values Θ for which we encountered surprising results.

We conjecture that the best strategy for the bounded guessing game is incompatible to the geometry of rational numbers.

9. NORMAL APPROXIMATION

We have seen that using binomial distribution for R_n yields surprising results but also makes our analysis very difficult. In this section we try to approximate the binomial distribution by normal distribution as is often done in biological setting. We approximate by a normal distribution with mean $n\psi$ and variance $n\Psi(1-\Psi)$ (i.e. the same mean and variance as the binomial distribution). This gives, exactly as in (6.8) and (6.9),

$$E_{normal}(\Omega_n) = E_{normal}\left(\frac{R_n}{n}\right) = \frac{E_{normal}(R_n)}{n} = \Psi$$

and

$$\mathrm{var}_{normal}(\Omega_n) = \mathrm{var}_{normal}\left(\frac{R_n}{n}\right) = \frac{\mathrm{var}_{normal}(R_n)}{n^2} = \frac{\Psi(1-\Psi)}{n}.$$

Denote $\sigma_n = \mathrm{var}_{normal}(\Omega_n)$. The fitness is then given by

$$\begin{aligned} F_{normal}(n, \Psi, \Theta) &= P(\Omega_n \leq \Theta) \\ &= \int_{-\infty}^{\Theta} \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(\frac{(x-\Psi)^2}{2\sigma_n^2}\right) \mathrm{d}x. \end{aligned}$$

Let $n < m$ so that $\sigma_n > \sigma_m$. Thus $\frac{\Theta-\Psi}{\sigma_m}\sigma_n + \Psi \lessgtr \Theta$ whenever $\Theta \gtrless \Psi$, and

$$\begin{aligned}\int_{-\infty}^{\Theta} \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(\frac{(x-\Psi)^2}{2\sigma_n^2}\right) \mathrm{d}x &= \int_{-\infty}^{\frac{\Theta-\Psi}{\sigma'}\sigma_n+\Psi} \frac{1}{\sigma'\sqrt{2\pi}} \exp\left(\frac{(x-\Psi)^2}{2\sigma'^2}\right) \mathrm{d}x \\ &\lessgtr \int_{-\infty}^{\Theta} \frac{1}{\sigma'\sqrt{2\pi}} \exp\left(\frac{(x-\Psi)^2}{2\sigma'^2}\right) \mathrm{d}x.\end{aligned}$$

Consequently

$$F(n,\Psi,\Theta) \lessgtr F(m,\Psi,\Theta), \quad \text{if } n < m \text{ and } \Theta \gtrless \Psi. \tag{9.13}$$

We just proved the following theorem:

Theorem 4. *Assuming the normal distribution for the number of incompatible drones the queen mated, the queen should mate with as many drones as possible if* $\Psi < \Theta$ *and with only* 1 *drone if* $\Psi > \Theta$.

This result is generally in agreement with the result for a smoother fitness function G, as well as with the Corollary 2. We can see that as long as the number of drones is made somehow fuzzy rather than exact (either directly through the formula for the fitness function, or indirectly by smoothening the underlying distribution), the queen has only two choices - either to mate with many (if the tolerance for incompatible drones is high relative to the real occurrence) or with one drone only.

10. Summary

We proved that on average, the proportion of incompatible drones queen mated does not depend on the number of drones she mated with; however, the variance of that proportion goes to 0 as the number of mates increases. We conclude that if there are less incompatible drones than the queen can tolerate, and if the queen can mate with many drones, then she should mate with as many as possible. Also, if there are more incompatible drones than the queen can tolerate, then mating with too many drones decreases the fitness to 0. However, in this case, we were unable to provide the number of drones the queen should mate to optimize the fitness.

The computer simulations suggest that if the queen cannot count (which is the case) and if there are more incompatible drones than she can tolerate, she should mate with 1 drone only.

Interestingly enough, if the queen could count, and her tolerance to incompatible drones were greater than roughly 1/7, then she should mate with more than 1 drone even if her tolerance is less than the real fraction of incompatible drones.

We also introduced, of independent interest,“guessing game” and its variant “bounded guessing game” which are equivalent to the biologically motivated problems. With the help of computer simulations, we conjectured that the best strategy for the bounded guessing game is related to the geometry of rational numbers.

References

[1] Andersson M (1994) Sexual Selection. Princeton University Press, Princeton, NJ.

[2] Baudry E, Solignac M, garnery L, Cornuet JM, Koeniger N (1998) Relatedness among honeybees (*Apis mellifera) of a drone congregation. Proceedings of the Royal Society of London B-Biological Sciences 265:2009-2014.*

[3] Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003) The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114:419-429.

[4] Boomsma JJ, Fjerdingstad EJ, Frydenberg J (1999) Multiple paternity, relatedness and genetic diversity in Acromyrmex leaf-cutter ants. Proceedings of the Royal Society of London Series B-Biological Sciences 266:249-254.

[5] Brown MJF, Schmid-Hempel P (2003) The evolution of female multiple mating in social hymenoptera. Evolution 57:2067-2081.

[6] Camazine S (1991) Self-organization pattern-formation on the combs of honey bee colonies. Behavioral Ecology and Sociobiology 28 (1): 61-76.

[7] Cole BJ (1983) Multiple Mating and the Evolution of Social-Behavior in the Hymenoptera. Behavioral Ecology and Sociobiology 12:191-201.

[8] Crozier RH, Fjerdingstad EJ (2001) Polyandry in social Hymenoptera - disunity in diversity? Annales Zoologici Fennici 38:267-285.

[9] Denny AJ, Franks NR, Powell S, Edwards KJ (2004) Exceptionally high levels of multiple mating in an army ant. Naturwissenschaften 91:396-399.

[10] Eishchen FA, Rothenbuhler WC, Kolincevic JM (1982) Length of life and dry weight of worker honeybees reared in colonies with different worker-larva ratios. Journal of Apicultural Research 21(1): 19-25.

[11] Foster KR, Ratnieks FLW (2001) Paternity, reproduction and conflict in vespine wasps: a model system for testing kin selection predictions. Behavioral Ecology and Sociobiology 50:1-8.

[12] Fukuda H, Ohtani T (1977) Survival and lifespan of drone honeybees. Res. Pop. Ecol. 19:51-68.

[13] Hasselman M, Beye M (2004) Signatures of selection among sex-determining alleles of the honey bee. Proceedings of the National Academy of Sciences of the United States of America 101 (14): 4888-4893.

[14] Kronauer DJC, Berghoff SM, Powell S, Denny AJ, Edwards KJ, Franks NR, Boomsma JJ (2006) A reassessment of the mating system characteristics of the army ant Eciton burchellii. Naturwissenschaften 93:402-406.

[15] Mackensen O (1951) Viability and sex determination in the honey bee (Apis mellifera L.). Genetics 36:500-509.

[16] Neumann P, Moritz RFA, van Praagh J (1999) Queen mating frequency in different types of honey bee mating apiaries. Journal of Apicultural Research 38:11-18.

[17] Oldroyd BP, Fewell JH (2007) Genetic diversity promotes homeostasis in insect colonies. Trends in Ecology & Evolution 22:408-413.

[18] Page RE (1980) The Evolution of Multiple Mating-Behavior by Honey Bee Queens Apis-Mellifera L. Genetics 96:263-273.

[19] Palmer KA, Oldroyd BP (2000) Evolution of multiple mating in the genus Apis. Apidologie 31:235-248.

[20] Sakagami SF, Fukuda H (1968) Life tables for worker honeybees. Research in Population Ecology 10:127-139.

[21] Schluns H, Moritz RFA, Neumann P, Kryger P, Koeniger G (2005) Multiple nuptial flights, sperm transfer and the evolution of extreme polyandry in honeybee queens. Animal Behaviour 70:125-131.

[22] Seeley TD, Tarpy DR (2007) Queen promiscuity lowers disease within honeybee colonies. Proceedings of the Royal Society B-Biological Sciences 274:67-72.

[23] Tarpy DR, Gilley DC, Seeley TD (2004) Levels of selection in a social insects: a review of conflict and cooperation during honey bee (Apis mellifera) queen replacement. Behavioral Ecology and Sociobiology 55:513-523.

[24] Tarpy DR, Page RE (2002) Sex determination and the evolution of polyandry in honey bees (Apis mellifera). Behavioral Ecology and Sociobiology 52:143-150.

[25] Tribe GD, Fletcher DJC (1977) Rate of the development of the workers of Apis mellifera adansonii L. In African bees: their taxonomy, biology, andeconomic use, ed. D.J.C. Fletcher, pp.115-119. Pretoria, Adimondia.

[26] Wattanachaiyingcharoen W, Oldroyd BP, Wongsiri S, Palmer K, Paar R (2003) A scientific note on the mating frequency of Apis dorsata. Apidologie 34:85-86.

[27] Wiernasz DC, Perroni CL, Cole BJ (2004) Polyandry and fitness in the western harvester ant, Pogonomyrmex occidentalis. Molecular Ecology 13:1601-1606.

[28] Winston ML (1987) The Biology of Honey Bees. First Harvard University Press paperback edition, 1991, p.55.

[29] Woyke J (1963) What happens to diploid drone larvae in a honey bee colony? J. Apic. Res. 2:73-75.

[30] Woyke J (1973) Reproductive organs of haploid and diploid drone honeybees. J. Apic. Res. 12:35-51.

Department of Mathematics and Statistics, University of North Carolina at Greensboro, NC27403, USA
E-mail address: `m_chhetr@uncg.edu`

Department of Statistics, Virginia Tech University, Blacksburg, VA, 24061, USA
E-mail address: `nels@vt.edu`

Department of Biology, University of North Carolina at Greensboro, NC27403, USA
E-mail address: `o_ruppel@uncg.edu`

Department of Mathematics and Statistics, University of North Carolina at Greensboro, NC27402, USA
E-mail address: `rychtar@uncg.edu`