ORIGINAL ARTICLE

THE ENFORCEMENT OF COOPERATION BY POLICING

Claire El Mouden,[1,2] Stuart A. West,[1] and Andy Gardner[1]

[1]*Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, United Kingdom*

[2]*E-mail: claire.elmouden@zoo.ox.ac.uk*

Policing is regarded as an important mechanism for maintaining cooperation in human and animal social groups. A simple model providing a theoretical overview of the coevolution of policing and cooperation has been analyzed by Frank (1995, 1996b, 2003, 2009), and this suggests that policing will evolve to fully suppress cheating within social groups when relatedness is low. Here, we relax some of the assumptions made by Frank, and investigate the consequences for policing and cooperation. First, we address the implicit assumption that the individual cost of investment into policing is reduced when selfishness dominates. We find that relaxing this assumption leads to policing being favored only at intermediate relatedness. Second, we address the assumption that policing fully recovers the loss of fitness incurred by the group owing to selfishness. We find that relaxing this assumption prohibits the evolution of full policing. Finally, we consider the impact of demography on the coevolution of policing and cooperation, in particular the role for kin competition to disfavor the evolution of policing, using both a heuristic "open" model and a "closed" island model. We find that large groups and increased kin competition disfavor policing, and that policing is maintained more readily than it invades. Policing may be harder to evolve than previously thought.

KEY WORDS: Island model, kin selection, repression of competition, scale of competition, tragedy of the commons.

Explaining cooperation is a major challenge for evolutionary biology. Natural selection favors those individuals who have the greatest relative fitness (Darwin 1859; Fisher 1930; Price 1970). However, all else being equal, cooperation reduces the relative fitness of the actor, as it involves providing a benefit to other individuals. The challenge is therefore to explain how cooperation can evolve by natural selection (Hamilton 1966; Sachs et al. 2004; Lehmann and Keller 2006; Lehmann et al. 2007; West et al. 2007a). In the social sciences, the problem of cooperation is famously encapsulated by the "tragedy of the commons," (Hardin 1968) which considers a group of herdsmen deciding how many animals to graze on a shared pasture. The benefit of adding an extra animal accrues directly to the individual herdsman while the cost of overgrazing is shared by all. Hence, each herdsman is favored to add further animals to the commons, resulting in severe overgrazing and the destruction of the shared resource. Finding solutions to avert this tragedy is of interest to evolutionary biologists, economists, and moral and political philosophers (Ostrom 1990; Binmore 1994, 1998; Frank 2003, 2009).

Two major ways for the tragedy of the commons to be averted are through kin-selected self-restraint or repression of competition (Frank 2003). The former mechanism works when relatives interact, as individuals that improve the reproductive success of those with which they share genes are more successful in transmitting copies of these genes to future generations (Hamilton 1963, 1964a; Maynard Smith 1964; Hamilton 1970). However, theory predicts that even groups of close relatives can foster significant internal conflict, and so there is still often scope for very costly competition within groups (Frank 1995, 1996b, 2003). Repression of competition is when there is a mechanism that reduces competition among the members of a group (Frank 2003). In contrast to kin-selected self-restraint, complete repression of competition is predicted to lead to harmonious groups because the individual can pursue the maximization of its inclusive fitness only by promoting the success of its group (Leigh 1971, 1977; Alexander 1987; Frank 2003; Gardner and Grafen 2009).

A potentially important means of repressing competition is by policing. The theory of policing has developed in two

directions: (1) general theory intended to capture how mechanisms that enforce an equitable distribution of shared group resources may evolve (e.g., Frank 1995, 1996b, 2003, 2009); and (2) theory developed specifically for eusocial insects (Ratnieks et al. 2006; Ratnieks and Helanterä 2009) such as honey bees where egg-laying workers are policed (their eggs are eaten) (Ratnieks 1988; Ratnieks and Visscher 1989). This work considers how mechanisms of kin-discrimination can lead to policing in haplodiploid societies. Here, we are concerned with the former approach, which aims for an illustrative overview of how the shared benefits of cooperation can drive the evolution of policing mechanisms to limit within-group conflict. Such policing is found at all scales of biological complexity: for example, intragenomic conflict is limited by fair meiosis (Leigh 1977); uncooperative symbionts can face sanctions from their host, such as in the legume–rhizobia mutualism (West et al. 2002a; Kiers et al. 2003; Kiers and Van Der Heijden 2006), the yucca–yucca moth mutualism (Pellmyr and Huth 1994; Pellmyr 2003), the fig tree–fig wasp mutualism (Jandér and Herre 2010) and the *Glochidon* tree–*Epicephala* moth mutualism (Goto et al. 2010); and "reproductive leveling" in human tribal groups is enforced via strict food sharing norms (Bowles 2006). Such general models highlight how comparable evolutionary mechanisms may result in cooperation evolving in very different contexts.

We develop the general theory of "group-benefits" driven policing (Frank 1995, 1996b), in three ways. First, we relax an implicit assumption in the original model that the individual cost of investment into policing is lower in more selfish groups. Second, we relax the assumption that the group cost of selfishness is totally recovered by policing. Third, we consider policing in a demographic context: the population processes that allow relatedness to build up (e.g., low dispersal) can often lead to intensified competition between relatives (Queller 1992; Taylor 1992; West et al. 2002c), but this has been neglected in previous models of policing. We consider the general impact of increased kin competition on the evolution of group-benefits driven policing, and provide an illustration using Wright's (1931) infinite island model.

Models and Analyses
FRANK'S MODEL OF SELFISHNESS AND POLICING
In this section, we introduce and rederive the results of the policing model developed by Frank (1995, 1996b, 2003, 2009), which forms the basis of our subsequent analyses. Frank (1995, 1996b) considered the evolution of a policing behavior that represses competition and limits the detrimental impact of within-group competition on the level of group resources, assuming that the fitness w of an individual is given by

$$w \propto \left(a' + (1 - a')\frac{z}{z'} - ca \right)(1 - (1 - a')z'), \qquad (1)$$

where a is the individual's investment into group policing; a' is the average level of policing by members of the group; z is the individual's investment into selfishness; z' is the group average level of selfishness; and c scales the individual cost of investment into policing (a summary of all notation is given in Table 1). Fitness is given by the individual's relative share (first multiplicative term) of total group resources (second multiplicative term). The individual's relative share of group resources is given by 1 in all policed encounters, and by her relative selfishness z/z' in all unpoliced encounters. The total group resources are reduced as unpoliced encounters and selfishness increase within the group. (A variant of the model considered by Frank (1996b), implements a personal cost of investment into policing alternatively as a third multiplicative fitness component, which does not qualitatively affect results.) Importantly, Frank's (1995) model assumes that investment into policing incurs a reduction in the individual's share of the group resources, and hence the absolute cost of policing is lower when the group has fewer resources (i.e., it is cheaper to police a more selfish group). We relax this assumption in the next section.

In the absence of policing ($a = a' = 0$), the model describes a basic tragedy of the commons scenario (Frank 1996a, 2009, 2010), where fitness is given by

Table 1. A summary of model notation.

| Symbol	Definition
a	Policing behavior
a'	Group average level of policing
z	Selfishness/competitiveness
z'	Group average level of selfishness/competitiveness
c	Cost of policing
b	Efficiency of policing at restoring group resources
r	Coefficient of relatedness ("whole-group" relatedness)
w	Darwinian fitness
R	"Others-only" relatedness
B	Fitness benefit to the recipient of a social behavior
C	Fitness cost to actor of a social behavior
n	Number of breeding spaces per group (group size)
F	Fecundity of an individual
F'	Mean fecundity of group
$\bar{F}$	Mean fecundity of population
s	Scale of competition
m	Dispersal rate
K	Number of offspring
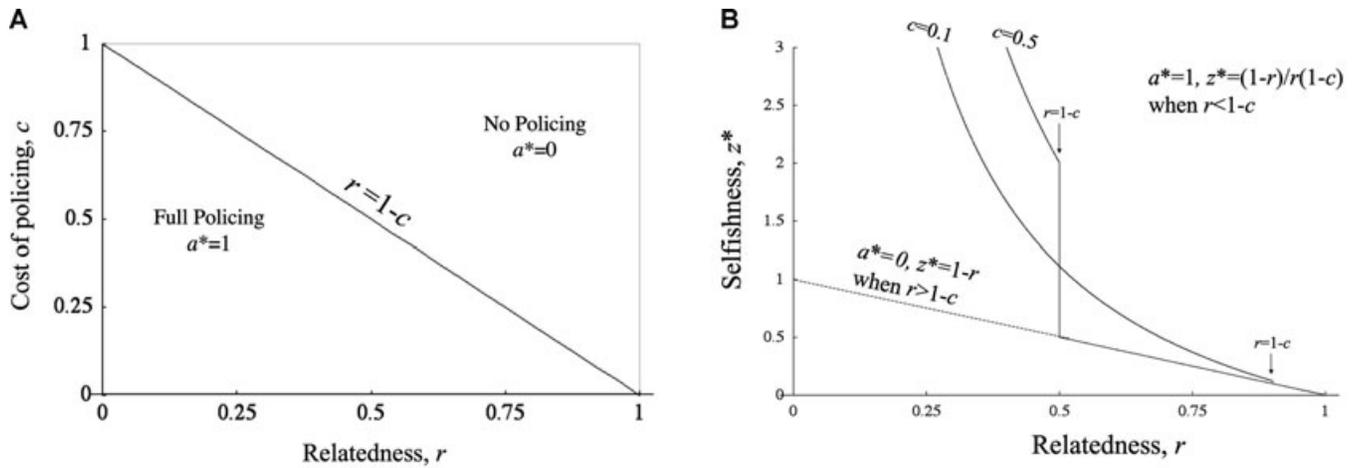
Figure 1. Frank's model of policing and selfishness. In Frank's policing model (Frank 1995; Figure 1A) full policing ($a^* = 1$), which results in a complete repression of within-group competition and a fair distribution of group resources, is predicted to invade whenever relatedness is low ($r < 1 - c$). When relatedness is high ($r > 1 - c$), no policing occurs ($a^* = 0$) and the model reduces to the tragedy of the commons and selfishness (Fig. 1B) is limited only by kin-selected self-restraint ($z^* = 1 - r$). Policing always favors increased selfishness ($dz^*/da > 0$, $z^* = (1 - r)/r(1 - c)$).

$$w \propto \frac{z}{z'}(1 - z'), \qquad (2)$$

where an individual's share of group resources increases with its relative investment in selfishness (z/z'), which favors the evolution of increased selfishness, and the tragedy occurs because a more selfish group has reduced resources ($1 - z'$). The model shows that one way the tragedy can be averted is when there is within-group relatedness, which favors some degree of self-restraint. If within-group relatedness is r, then the evolutionarily stable (ESS; Maynard Smith and Price 1973) level of selfishness is $z^* = 1 - r$ (Frank 1995, 1996b).

Frank (1995) found that relatively high relatedness disfavors policing ($a^* = 0$ when $r > 1 - c$; Fig. 1A). This is because kin-selected self-restraint can maintain relatively high cooperation in high-relatedness groups, so that the benefits of policing do not outweigh its costs. Note, however, that even in groups with considerable relatedness, there is scope for significant selfishness. For example, in a large group of full siblings, relatedness is approximately one-half, so the ESS level of selfishness is also equal to one-half, and hence the group's fitness is only half as great as it could be if selfishness were completely abolished. Conversely, a relatively low within-group relatedness favors full investment into policing ($a^* = 1$ when $r < 1 - c$; Fig. 1A). A further result, not emphasized by Frank (1995), is that policing results in increased selfishness $z^* > 1 - r$; Fig. 1B). Hence, Frank's model of repression of competition by policing works by recovering the group cost of selfishness and not by reducing selfishness itself.

Frank (1995, 1996b) considered the evolutionary dynamics and statics of policing and selfishness, but did not show how policing fits into the standard classification of social behaviors provided

by Hamilton (1964b; Table 2). The classification depends on the others-only relatedness of group mates (R; the expected relatedness of two individuals sampled at random from the same group without replacement; Pepper 2000). Details of how the classification is determined are given in the Appendix. We find that

> **Result 1** – In Frank's (1995) model, policing may be altruistic or mutually beneficial. Full policing is altruistic when others-only relatedness is high ($R > [n(1 - c) - 1]/[(n - 1)(1 + cn)]$) and mutually beneficial when others-only relatedness is low ($R < [n(1 - c) - 1]/[(n - 1)(1 + cn)]$).

THE INDIVIDUAL COST OF POLICING

Owing to multiplicativity of terms, Frank's (1995, 1996b) model of policing implicitly assumes that the magnitude of the individual cost of investment into policing increases with group resources (i.e., $\partial w/\partial a = -c(1 - (1 - a')z')$, where $1 - (1 - a')z'$ is the total group resources). In other words, policing gets cheaper to maintain as groups become more selfish. However, while this could apply

Table 2. Classification of social behavior. Hamilton's rule states that a behavior or trait will be favored by selection, when $RB - C > 0$; where C is the fitness cost to the actor, B is the fitness benefit to the recipient weighted by R, their others-only relatedness. A behavior is termed altruistic if $B > 0$ and $C > 0$, mutually beneficial if $B > 0$ and $C < 0$, selfish if $B < 0$ and $C < 0$, and spiteful if $B < 0$ and $C > 0$ (Hamilton 1964b; West et al. 2007b).

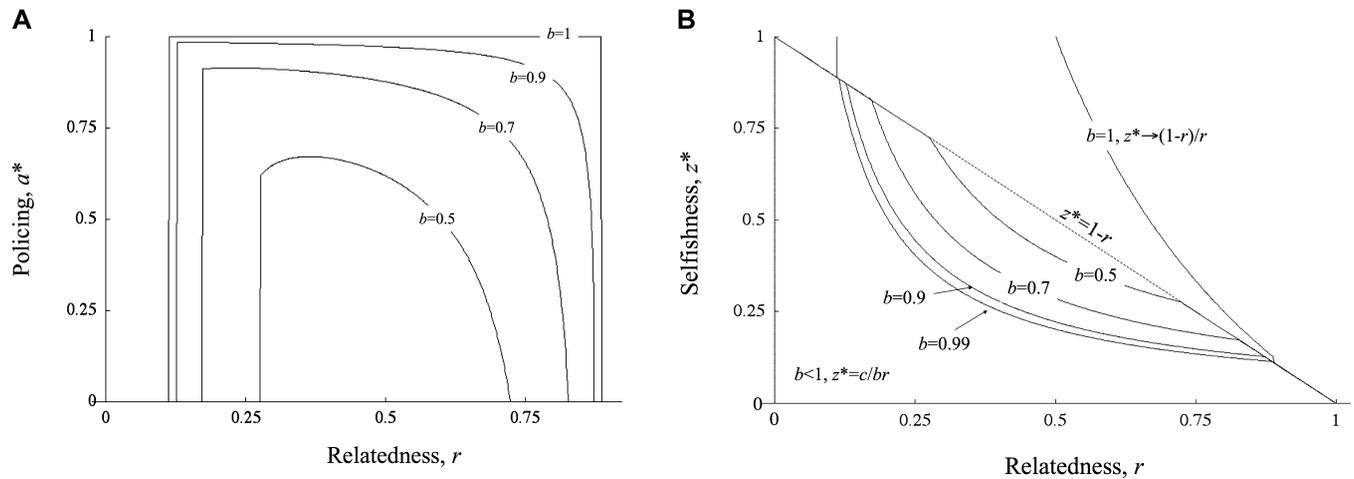| Effect on actor ($-C$)	Effect on recipient (B)	
	+	−
+	Mutual benefit	Selfishness
−	Altruism	Spite

Figure 2. Modified individual and group cost models. In Frank's model, the cost of policing was proportional to the value of group resources so high levels of selfishness were favored as they allowed vanishingly cheap policing. When the cost is adjusted to be proportional only to the level of policing an individual performs (i.e., $\partial w/\partial a = -c$, we assume here $c = 0.1$; Fig. 2A) policing is favored to invade only at intermediate relatedness (when $r(1 - r) > c/b$). If the efficiency with which a policing behavior restores group resources is imperfect ($b < 1$) intermediate levels of policing are favored (Fig. 2A) and when it invades, it always promotes self-restraint ($dz^*/da < 0$, $z^* = c/br$; Fig. 2B). This is in contrast to perfect policing ($b = 1$, $a = 1$), which favors increased selfishness ($z^* \to (1 - r)/r$; Fig. 2B). When policing is not stable ($a^* = 0$ is stable when $r(1 - r) < c/b$), the model reduces to the tragedy of the commons ($z^* = 1 - r$). As the efficiency of a policing behavior at restoring group resources b is reduced, the ESS level of policing falls ($da^*/db < 0$; see Appendix).

in situations where the cost of policing is primarily manifested as a reduction in an individual's ability to compete for group resources, this seems unlikely to hold in general. In this section, we assume instead that policing requires a fixed investment of time or energy. The true relationship between the cost of policing and selfishness will depend on the biological details; our aim here is to illustrate the importance of such details and provide a base upon which more complex models that consider specific biological cases may be developed.

We make the marginal cost of policing a fixed proportion of group fitness ($\partial w/\partial a = -c$), which we implement using a modified fitness function

$$w \propto \left(a' + (1 - a')\frac{z}{z'}\right)(1 - (1 - a')z') - ca. \qquad (3)$$

In the Appendix, we show that this leads to the following predictions:

Result 2 – Full policing ($a^* = 1$) is only favored at intermediate relatedness (specifically, when $r(1 - r) > c$; Fig. 2A). At the extremes of relatedness ($r(1 - r) < c$), policing cannot evolve and the model recovers the tragedy of the commons result ($a^* = 0$ and $z^* = 1 - r$). This is in contrast to Frank's (1995, 1996b) model, in which policing is favored at arbitrarily low relatedness.

Result 3 – Full policing leads to greater selfishness ($z^* = (1 - r)/r$; Fig. 1B) but is less than is predicted to occur for full policing in Frank's (1995, 1996b) model.

THE GROUP COST OF SELFISHNESS

Frank's (1995, 1996b, 2003) models assumed that policing fully recovers the loss of group fitness due to selfishness. This seems appropriate if policing removes the opportunity for individuals to behave selfishly (as in fair meiosis; (Leigh 1977). However, if policing simply works to remove the fitness benefits of selfishness, without directly suppressing selfishness itself (as in the eating of worker-laid eggs in honey bees, which does not directly prevent worker egg-laying but removes the benefits of doing this; Ratnieks et al. 2006), then we would expect selfishness to continue to incur some group-level cost even in the presence of full policing. We now develop the model described in the previous section to allow for a less than perfect efficiency of policing, with only a proportion b of the group fitness lost to selfishness being recovered by policing. In particular, we write

$$w \propto \left(a' + (1 - a')\frac{z}{z'}\right)(1 - (1 - ba')z') - ca. \qquad (4)$$

We find that this adjustment impacts upon the model behavior in several ways (see Appendix for details)

Result 4 – Full policing is never stable ($a^* < 1$; a full expression for a^* is given in the Appendix). This is in contrast to Frank's (1995, 1996b) model, in which full policing may be stable ($a^* = 1$), and an intermediate level of policing is never stable.

Result 5 – Reduced efficiency of policing (lower b) reduces the scope for policing to invade ($a^* > 0$ only if $r(1 - r) > c/b$, see

Appendix for full expression of a^*) and leads to a lower ESS level of policing when it does invade ($da^*/db > 0$; Fig. 2A).

Result 6 – Policing is more readily maintained by selection than it initially evolves. The condition for invasion of policing (instability of $a = 0$) is more stringent that the condition for an internal stable equilibrium to exist ($0 < a^* < 1$; Fig. 4).

Result 7 – Policing promotes self-restraint ($z^* < 1-r$; Fig. 2B). When there is policing, selfishness is $z^* = c/rb$ which, in the region where policing can invade from rarity ($r(1 - r) > c/b$), is lower than that which is obtained in the absence of policing ($z^* = 1 - r$)). This is in contrast to Frank's (1995, 1996b) model, where policing always favors increased selfishness ($z^* > 1 - r$; Fig. 1B)

Result 8 – Policing is an altruistic behavior. This is in contrast to Frank's (1995, 1996b) model, in which policing can be either altruistic or mutually beneficial (see Result 1).

DEMOGRAPHY AND POLICING—ILLUSTRATIVE OVERVIEW

Frank's (1995, 1996b) models consider the impact of whole-group relatedness (r) upon the coevolution of policing and cooperation, and our developments of this basic model reveal that an intermediate level of relatedness is crucial for policing to evolve. However, the ecological and demographic processes that lead to relatedness within groups can often also lead to intensified competition between relatives (reviewed by Queller 1992; West et al. 2002c), which has not been explicitly considered in Frank's models. For example, if individuals do not disperse far from where they were born (population viscosity, Hamilton 1964b), then relatives must compete more intensely for space and resources.

We extend the model developed in the previous section, by making explicit the impact of policing and selfishness upon fecundity, and by making explicit the impact of personal fecundity and the fecundity of competitors upon Darwinian fitness. We employ a "scale of competition parameter" (s; Frank 1998), which measures the extent to which an increase in an individual's fecundity leads to intensified competition within its social group. This approach has been useful for both theoretical (West et al. 2002c; West and Buckling 2003; Gardner and West 2004b, 2006; Gardner et al. 2009; Lively 2009; Gardner 2010) and empirical (West et al. 2001; Griffin et al. 2004; West et al. 2006; Kümmerli et al. 2009) studies of social evolution. Following the approach taken by those earlier studies, we define the fitness of an individual as

$$w = \frac{F}{sF' + (1-s)\bar{F}}, \qquad (5)$$

where F is the relative fecundity of the individual, and is given by

$$F = \left(a' + (1-a')\frac{z}{z'}\right)(1 - (1 - ba')z') - ca; \qquad (6)$$

F' is the average fecundity of the individuals in its social group,

and is given by

$$F' = 1 - (1 - ba')z' - ca'; \qquad (7)$$

and $\bar{F}$ is the average fecundity of individuals over the whole population, and is given by

$$\bar{F} = 1 - (1 - b\bar{a})\bar{z} - c\bar{a}. \qquad (8)$$

Thus the fitness of an individual is determined by their fecundity (F) relative to the fecundity of their competitors, a proportion s of which are locals (group mates, who have average fecundity F') and a proportion $1 - s$ of which are random individuals drawn from the population as a whole (having average fecundity $\bar{F}$). In the Appendix, we derive the following results:

Result 9 – Local competition reduces the scope for policing to invade ($a^* > 0$ when $r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$, which is increasingly stringent as s increases; for a^*, see appendix). Increased local competition both narrows the range of relatedness values that favor the invasion of policing ($d[r_{max} - r_{min}]/ds < 0$ where r_{max} and r_{min} are the maximum and minimum relatedness permitting policing to invade from rarity) and shifts this range to higher levels of relatedness ($dr_{max}/ds > 0$, $dr_{min}/ds > 0$; Fig. 3A). When competition is entirely local ($s = 1$) policing is never favored. Policing is still maintained more readily than it invades (Fig. 4).

Result 10 – Increased local competition leads to higher selfishness ($dz^*/ds > 0$). In the absence of policing $z^* = (1 - r)/(1-rs)$ which rises from $z^* = 1 - r$ when $s = 0$ to $z^* = 1$ when $s = 1$. As before, policing promotes self-restraint ($z^* = c(1 - rs)/br(1 - s)$) in the region where it can invade from rarity (when $r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$).

Queller (1994) suggested that the effects of local competition can be easily accounted for by redefining relatedness so that it measures genetic similarity of two individuals with respect to their "economic neighborhood" (the arena in which they compete) rather than the with respect to the population as a whole. The condition for policing $r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$ may be rewritten as $\rho(1 - \rho) > c/b$ by substituting whole-group relatedness, r, with Queller's relatedness $\rho = (r - sr)/(1 - sr)$ (see also Gardner and West 2006). Queller's relatedness captures both the genetic structure of a population (relatedness, r) and the effect of changing the scale of competition (s), with Queller's relatedness being an increasing function of the former ($\partial\rho/\partial r > 0$) and a decreasing function of the latter ($\partial\rho/\partial s < 0$). Results 9 and 10 may be derived from the basic model results (Results 2, 4–7) by substituting "r" with "ρ." If competition is entirely local, regardless of the genetic relatedness, there is no incentive to perform cooperative behaviors as all group members are valued equally (i.e., when $s = 1$, $\rho = 0$). This is why the effect of kin-selected self-restraint on selfishness disappears as competition becomes localized, which then confines policing to higher levels of relatedness.
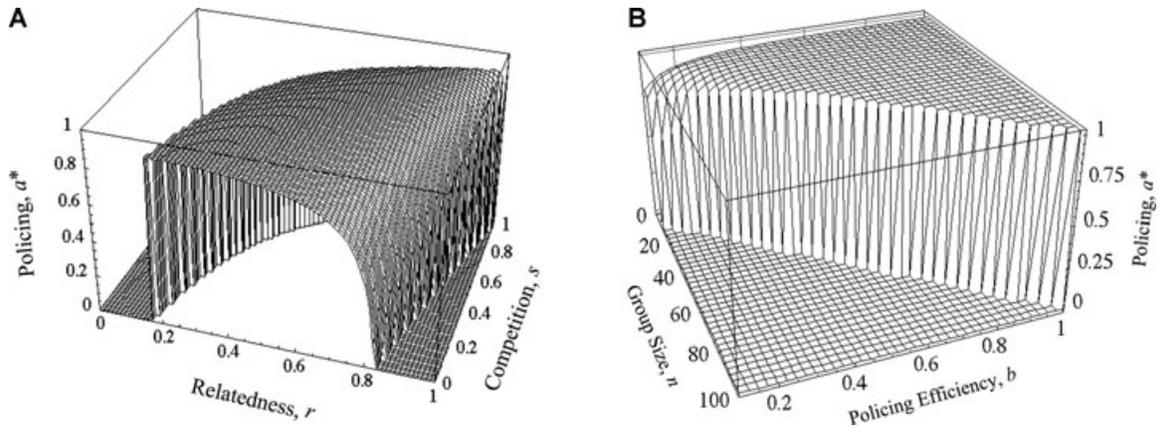
Figure 3. Policing and demography. Increased local competition reduces the scope for policing ($a^* > 0$ when $r(1 − r)[(1 − s)/(1 − rs)^2] > c/b$). In the open model (Fig. 3A, where we assume $c = 0.01$, $b = 0.7$), as the scale of competition (s) increases, policing is restricted to higher, narrower levels of relatedness. When $s = 0$, policing is equal to the line for $b = 0.7$ in Fig. 2A. In the Island model (Fig. 3B, where we assume $c = 0.1$), which considers a population subdivided into groups of size n with migration occurring between them, we find that policing invades less readily in larger groups ($a^* > 0$ when $(n − 1)/n^2 > c/b$).
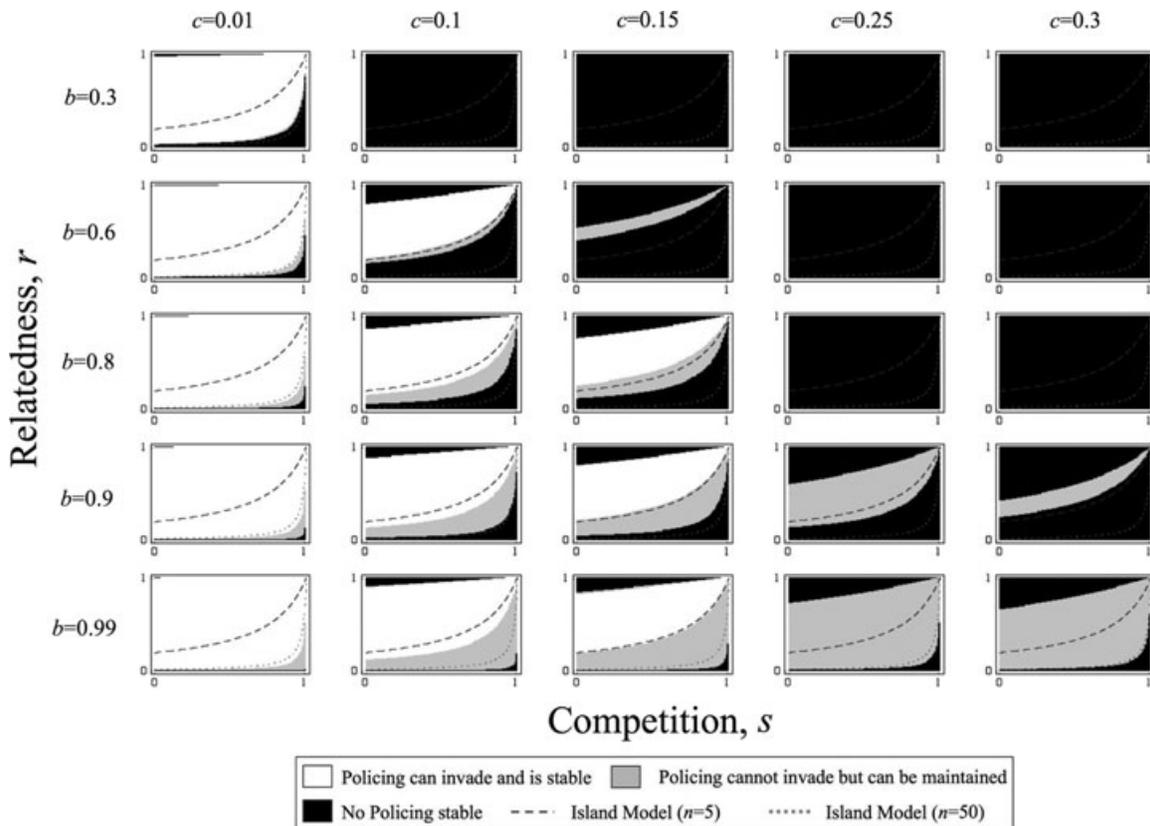

Figure 4. Policing and demography—stability analysis. A numerical investigation (following the methodology of Otto and Day 2007, pp. 237–242) shows that the condition for maintenance of policing is less stringent than the condition for it to increase from rarity. The gray zone indicates that both policing ($a^* > 0$) and no policing ($a^* = 0$) are stable (i.e., policing can be maintained but cannot invade); white indicates that zero policing is unstable (can invade and can be maintained, which is when $c/b < r(1 − r)[(1 − s)/(1 − rs)^2]$); black indicates that policing cannot invade nor be maintained ($a^* = 0$). The island-model stability results can be calculated from the openmodel results as $s = (1 − m)^2$ and $r = 1/[n − (n − 1)s]$. They illustrate that once present, policing can be maintained in much larger groups (plotted for $n = 5$ and $n = 50$).

DEMOGRAPHY AND POLICING—ISLAND MODEL

In the previous section, we provided an illustrative overview of how intensified kin competition impacts the coevolution of policing and cooperation. This revealed that, all else being equal, increased competition reduces the scope for policing and self-restraint to evolve. In particular, we examined the impact of increased kin competition (s) under the assumption of fixed relatedness (r). However, the same demographic processes may determine the degree of relatedness and kin competition, so it can be misleading to consider the effects of changing the value of either of these quantities in isolation. Here, we explicitly consider the codependency of relatedness and kin competition in a specific infinite island model setting (Wright 1931). The model assumes that an infinite population of asexual haploid individuals compete for resources in groups of size n. The relative fecundity of an individual is determined by its success in social interactions within the group (as in eqs. 6–8). Individuals then have a very large number of offspring (K) proportional to their fecundity, of which a proportion m migrate to randomly chosen groups and a proportion $1 - m$ remain in the natal group. We assume that parents die at each generation and that dispersing individuals will not find relatives in their new group. In the next generation, native and immigrant individuals compete in every group for the n breeding spots. The fitness of an individual is determined by its success relative to others in its group. In the Appendix, we show that

> **Result 11** – The migration rate m has no net effect on the evolution of policing ($da^*/dm = dz^*/dm = 0$). Decreasing the migration rate both increases relatedness ($r = 1/(n - (n - 1)(1 - m)^2$, so $dr/dm < 0$) and also increases local competition ($s = (1 - m)^2$, $ds/dm < 0$). However these effects have no net impact on Queller's relatedness ρ, which is the relatedness of an individual to its social partners relative to its competitors, (substituting the island model values for r and s into ρ gives an effective relatedness of $\rho = 1/n$ which is not a function of m).

> **Result 12** – Increasing group size disfavors the invasion of policing (policing can increase from rarity when $(n - 1)/n^2 > c/b$, in large groups this is approximately when $b/c > n$, for a^* see Appendix; Fig. 3B). The larger the group, the lower the whole-group relatedness (which falls from $r = 1$ when the group size $n = 1$ to $r \to 0$ as $n \to \infty$) and the higher the selfishness. Policing is maintained in larger groups more readily than it invades in them (Fig. 4). In the absence of policing, selfishness is stable at $z^* = (n - 1)/n$ and as before, in the region in which policing can invade, it always promotes self-restraint, given by $z^* = cn/b$.

Discussion

Group-benefits driven policing—enforcement of fairness in the distribution of resources among group mates—has been regarded as a potentially important mechanism for the evolution of coopera-tion and the major transitions in evolution (Leigh 1977; Alexander 1987; Frank 2003; Gardner and Grafen 2009). In this article, we have examined Frank's (1995, 1996b) model of policing, which was developed to provide a general overview of group-benefits driven policing theory. The basic predictions of his model were that policing evolves most readily when within-group relatedness is low, that this leads to an increase in selfishness and that, when it does evolve, policing will completely abolish competition among group mates (Fig. 1). However, his model implicitly assumes that the individual cost of investment into policing falls as group selfishness increases, and also that policing totally recovers any loss of group fitness due to selfishness—even though individuals continue to invest in selfishness. We have examined the conse-quences of relaxing these assumptions and find these basic results are changed substantially. Specifically, we find that policing does not evolve when relatedness is low (Fig. 2A), favors a reduction in individual selfishness (Fig. 2B), cannot repress all within-group conflict (i.e., $a^* < 1$; Fig. 2A), is easier to maintain at lower relatedness than to initially evolve (Fig. 4), and is altruistic.

We also examined the consequences of placing group-benefits policing in its proper demographic context. We have shown that within-group relatedness is crucial for policing to be favored, and yet the demographic processes that can generate within-group relatedness may also lead to intensified kin com-petition, which has not been considered explicitly in previous models. We found that, in general, an increased competition for resources within groups leads to a reduction in policing. Fur-ther, in the context of Wright's (1931) infinite island model, we found that the relatedness-inflating effects of low migration ex-actly cancel the associated increase in local competition, so the evolution of policing is solely dependent upon group size, with larger groups disfavoring policing (Fig. 3B). In short, all of our model developments reduce the scope for policing to invade from rarity. However, our results also show that policing is easier to maintain than it is to evolve, specifically that it is stable at lower relatedness (i.e., in larger groups; Fig. 4). This result mirrors previous work on the evolution of punishment and cooperation which showed that punishment is maintained more readily than it initially evolves (Gardner and West 2004a).

We found that group-benefits policing requires intermedi-ate relatedness within groups to be favored by natural selection. We emphasize that this refers to the "whole-group" relatedness, which includes the relatedness of the individual to itself, and not the "others-only" relatedness, which is the relatedness be-tween social partners (Pepper 2000). Thus, the model could be applied to study policing in interspecific mutualisms, where a host individual is not genetically related to its symbiont(s), but nevertheless could reap the benefits of policing its social part-ners (mediated through nonzero whole-group relatedness). Our model predicts policing evolves most readily when whole-group

relatedness is one-half, which is obtained when unrelated partners are in a 1:1 relationship or when one individual forms a symbiosis with a clonal group. However, our model is not intended to fully capture the biology of mutualisms. Importantly, it treats all individuals as equivalent, incurring the same costs and benefits for the same investments into policing and selfishness, whereas real-world mutualisms bring nonequivalent individuals from different species together in a complementary relationship. Development of this basic policing model to explicitly incorporate the biology particular to mutualisms represents an important direction for future research (Murray 1985; West et al. 2002a,b; Yu et al. 2004; Foster and Wenseleers 2006).

Group-benefits driven policing may have been important in human evolution (Bowles 2006). As policing is maintained more readily than it invades, it could have evolved at a time when humans lived in small tightly knit groups and then been retained as much larger societies developed, particularly if social norms allowed the cost of policing to be reduced. To investigate the evolution of social behaviors in humans, the effect of both population structure and cultural transmission (if they are culturally inherited) on their evolution must be understood. Whilst the importance of cultural transmission in enhancing/limiting selection for cooperation has received considerable attention (Boyd and Richerson 1985; Henrich and Boyd 1998; Henrich and Gil-White 2001; Richerson 2005; Lehmann and Feldman 2008; Lehmann et al. 2008), the importance of population structure has not. For example, budding dispersal, where groups migrate to new territories (the "tribe-splitting" model of Haldane 1932) may be a better description of human population movements than the basic island model used here (which assumes individuals disperse independently). Both theoretical and empirical work (Gardner and West 2006; Lehmann et al. 2006; Kümmerli et al. 2009) show that budding dispersal favors the evolution of cooperation. This illustrates the importance of considering whether models capture the details of human population structure when investigating the evolution of behaviors such as policing. Extending the island model (Wright 1931; Taylor 1992) to study policing in a demographic context that more accurately reflects the structure of human Palaeolithic society is an important task for the future.

In this article, we have focused upon models of policing that work owing to group benefits (following Leigh 1977; Alexander 1987; Frank 2003). However, group benefits are not necessary for policing to be favored and for policing to favor cooperation. Perhaps the best understood policing system, from a theoretical and empirical standpoint, is that of honey bees and other hymenoptera. Here, workers prevent the rearing of eggs laid by other workers by eating them upon discovery, and this has been shown to reduce the extent to which workers do lay eggs (Wenseleers et al. 2004a,b). The selective force favoring policing in such cases is the nepotistic interests of the egg-eating workers, who value queen-laid unfertilized eggs above worker-laid unfertilized eggs, because they are more genetically related to the former than to the latter (Ratnieks 1988; Ratnieks and Visscher 1989; Wenseleers et al. 2005; Wenseleers and Ratnieks 2006; Ratnieks and Helanterä 2009). Consequently, this egg-eating behavior can evolve even when it does not provide a colony-level benefit (Wenseleers et al. 2004a,b). However, such policing can also provide colony-level benefits, which can maintain policing even in cases in which there are not relatedness asymmetries (Martin et al. 2002).

How do our results relate to our understanding of the major evolutionary transitions? A major transition can occur when individuality shifts from a lower to higher level (Buss 1987; Maynard Smith and Szathmary 1995; Queller 2000). For such transitions to be possible, internal conflict must be resolved (Frank 2003). This can occur owing to clonality of individuals within social groups, or if there is a total repression of within-group competition (Buss 1987; Maynard Smith and Szathmary 1995; Gardner and Grafen 2009). Policing has been suggested as a potentially important factor in driving major transitions (Frank 1995, 1996a,b, 2003; Ratnieks and Helanterä 2009). We suggest caution, as this may depend upon whether policing evolves via group-benefits or kin discrimination, and as the extent to which individuals can discover other ways to pursue their selfish interests. This suggestion is consistent with recent work on social insects, which suggests lifetime monogamy (which leads to high within-colony relatedness) was an essential step in the evolution of eusociality (Boomsma 2007, 2009; Hughes et al. 2008). It appears that worker policing was only important after eusociality became irreversible as a means to limit within-colony reproductive conflicts that emerged as queens began mate multiply (Ratnieks et al. 2006; Ratnieks and Wenseleers 2008). More generally, this emphasizes how the origin versus the maintenance evolution of cooperation behaviors can involve different selective forces (West et al. 2007a).

To conclude, we have shown that policing mechanisms that enforce the fair sharing of group resources may be harder to evolve than previously thought. We show that contrary to previous suggestions, policing can be disfavored at low relatedness. Furthermore policing may not result in a complete repression of within-group competition. This suggests group-benefits driven policing may be of only limited importance in major transitions. Major tasks for the future include the study of policing in different demographies, particularly with reference to humans and understanding the role of policing in intraspecific mutualisms.

LITERATURE CITED

Alexander, R. D. 1987. The biology of moral systems. Aldine de Gruyter, New York.

Binmore, K. 1994. Game theory and the social contract Volume 1: playing fair. MIT Press, Cambridge, MA.

———. 1998. Game theory and the social contract Volume 2: just playing. MIT Press, Cambridge, MA.

Boomsma, J. J. 2007. Kin selection versus sexual selection: why the ends do not meet. Curr. Biol. 17:R673–R683.

———. 2009. Lifetime monogamy and the evolution of eusociality. Proc. R. Soc. Lond. B 364:3191–3207.

Bowles, S. 2006. Group competition, reproductive leveling, and the evolution of human altruism. Science 314:1569–1572.

Boyd, R., and P. J. Richerson. 1985. Culture and the evolutionary process. Univ. of Chicago Press, Chicago.

———. 2005. The origin and evolution of cultures. Oxford Univ. Press, Oxford.

Buss, L. 1987. The Origin of Individuality. Princeton Univ. Press, Princeton, NJ.

Darwin, C. 1859. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. John Murray, London, UK.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Foster, K. R., and T. Wenseleers. 2006. A general model for the evolution of mutualisms. J. Evol. Biol. 19:1283–1293.

Frank, S. A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. Nature 377:520–522.

———. 1996a. Models of parasite virulence. Q. Rev. Biol. 71:37–78.

———. 1996b. Policing and group cohesion when resources vary. Anim. Behav. 52:1163–1169.

———. 1998. Foundations of social evolution, Princeton Univ. Press, Princeton, NJ.

———. 2003. Repression of competition and the evolution of cooperation. Evolution 57:693–705.

———. 2009. Evolutionary foundations of cooperation and group cohesion. Pp. 3–40 *in* S. A. Levin ed. Games, groups, and the global good. Springer-Verlag, Berlin.

———. 2010. Demography and the tragedy of the commons. J. Evol. Biol. 23:32–39.

Gardner, A. 2010. Sex-biased dispersal of adults mediates the evolution of altruism among juveniles. J. Theor. Biol. 262:339–345.

Gardner, A., and S. A. West. 2004a. Cooperation and punishment, especially in humans. Am. Nat. 164:753–764.

———. 2004b. Spite and the scale of competition. J. Evol. Biol. 17:1195–1203.

———. 2006. Demography, altruism, and the benefits of budding. J. Evol. Biol. 19:1707–1716.

Gardner, A., and A. Grafen. 2009. Capturing the superorganism: a formal theory of group adaptation. J. Evol. Biol. 22:659–671.

Gardner, A., A. Arce, and J. Alpedrinha. 2009. Budding dispersal and the sex ratio. J. Evol. Biol. 22:1036–1045.

Goto, R., T. Okamoto, E. T. Kiers, A. Kawakita, and M. Kato. 2010. Selective flower abortion maintains moth cooperation in a newly discovered pollination mutualism. Ecology Letters 13:321–329.

Griffin, A. S., S. A. West, and A. Buckling. 2004. Cooperation and competition in pathogenic bacteria. Nature 430:1024–1027.

Haldane, J. B. S. 1932. The causes of evolution. Longmans. New York.

Hamilton, W. D. 1963. The evolution of altruistic behaviour. Am. Nat. 97:354–356.

———. 1964a. The genetical evolution of social behaviour I. J. Theor. Biol. 7:1–16.

———. 1964b. The genetical evolution of social behaviour, I & II. J. Theor. Biol. 7:1–52.

———. 1966. The moulding of senescence by natural selection. J. Theor. Biol. 12:12–45.

———. 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228:1218–1220.

Hardin, G. 1968. The tragedy of the commons. Science 162:1243–1248.

Henrich, J., and R. Boyd. 1998. The evolution of conformist transmission and the emergence of between-group differences. Evol. Hum. Behav. 19:215–241.

Henrich, J., and F. J. Gil-White. 2001. The evolution of prestige freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evol. Hum. Behav. 22:165–196.

Hughes, W. O. H., B. P. Oldroyd, M. Beekman, and F. L. W. Ratnieks. 2008. Ancestral monogamy shows kin selection is the key to the evolution of eusociality. Science 320:1213–1216.

Jandér, C. K., and E. A. Herre. 2010. Host sanctions and pollinator cheating in the fig treefig-wasp mutualism. Proc. R. Soc. B. doi: 10.1098/rspb.2009.2157. *In press*.

Kiers, E. T., and M. G. A. Van Der Heijden. 2006. Mutualistic stability in the arbuscular mycorrhizal symbiosis: exploring hypotheses of evolutionary cooperation. Ecology 87:1627–1636.

Kiers, E. T., R. A. Rousseau, S. A. West, and R. F. Denison. 2003. Host sanctions and the legume-rhizobium mutualism. Nature 425:78–81.

Kümmerli, R., A. Gardner, S. A. West, and A. S. Griffin. 2009. Limited dispersal, budding dispersal and cooperation: an experimental study. Evolution 63:939–949.

Lehmann, L., and M. W. Feldman. 2008. The co-evolution of culturally inherited altruistic helping and cultural transmission under random group formation. Theor. Popul. Biol. 73:506–516.

Lehmann, L., and L. Keller. 2006. The evolution of cooperation and altruism. A general framework and classification of models. J. Evol. Biol. 19:1365–1378.

Lehmann, L., N. Perrin, and F. Rousset. 2006. Population demography and the evolution of helping behaviours. Evolution 60:1137–1151.

Lehmann, L., L. Keller, S. A. West, and D. Roze. 2007. Group selection and kin selection. Two concepts but one process. Proc. Natl. Acad. Sci. USA 104:6736–6739.

Lehmann, L., M. W. Feldman, and K. R. Foster. 2008. Cultural transmission can inhibit the evolution of altruistic harming. Am. Nat. 172:12–24.

Leigh, E. G. 1971. Adaptation and diversity. Cooper, San Francisco.

———. 1977. How does selection reconcile individual advantage with the good of the group? Proc. Nat. Acad. Sci. USA 74:4542–4546.

Lively, C. 2009. Local host competition in the evolution of virulence. J. Evol. Biol. 22:1268–1274.

Martin, S. J., M. Beekman, T. C. Wossler, and F. L. W. Ratnieks. 2002. Parasitic worker honey bees, *Apis mellifera capensis*, evade policing. Nature 415:163–165.

Maynard Smith, J. 1964. Group selection and kin selection. Nature 201:1145–1147.

Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. Nature 246:15–18.

Maynard Smith, J., and E. Szathmary. 1995. The major transitions in evolution. W.H. Freeman, Oxford.

Murray, M. G. 1985. Figs (*Ficus* spp.) and fig wasps (*Chalcidoidea*, *Agaonidae*): hypotheses for an ancient symbiosis. Biol. J. Linn. Soc. 26:69–81.

Ostrom, E. 1990. Governing the commons: the evolution of institutions for collective action. Cambridge Univ. Press, New York.

Otto, S. P., and T. Day. 2007. A biologist's guide to mathematical modeling in ecology and evolution. Princeton Univ. Press, Princeton, NJ.

Pellmyr, O. 2003. Yuccas, yucca moths, and coevolution: a review. Ann. Mo. Bot. Gard. 90:35–55.

Pellmyr, O., and C. J. Huth. 1994. Evolutionary stability of mutualism between yuccas and yucca moths. Nature 372:257–260.

Pepper, J. W. 2000. Relatedness in trait group models of social evolution. J. Theor. Biol. 206:355–368.

Price, G. R. 1970. Selection and covariance. Nature 227:520–521.

Queller, D. C. 1992. Does population viscosity promote kin selection? Trend. Ecol. Evol. 7:322–324.

———. 1994. Genetic relatedness in viscous populations. Evol. Ecol. 8:70–73.

———. 2000. Relatedness and the fraternal major transitions. Phil. Trans. R. Soc. Lond. B 355:1647–1655.

Ratnieks, F. L. W. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. Am. Nat. 132:217–236.

Ratnieks, F. L. W. H., and H. Helanterä. 2009. The evolution of extreme altruism and inequality in insect societies. Proc. R. Soc. Lond. B 364:3169–3179.

Ratnieks, F. L. W., and P. K. Visscher. 1989. Worker policing in the honeybee. Nature 342:796–797.

Ratnieks, F. L. W., and T. Wenseleers. 2008. Altruism in insect societies and beyond: voluntary or enforced? Trends Ecol. Evol. 23:45–52.

Ratnieks, F. L. W., K. R. Foster, and T. Wenseleers. 2006. Conflict resolution in insect societies. Annu. Rev. Entomol. 51:581–608.

Sachs, J. L., U. G. Mueller, T. P. Wilcox, and J. J. Bull. 2004. The evolution of cooperation. Q. Rev. Biol. 79:135–160.

Taylor, P. D. 1992. Altruism in viscous populations—an inclusive fitness model. Evol. Ecol. 6:352–356.

———. 1996. The selection differential in quantitative genetics and ESS models. Evolution 50:2106–2110.

Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. J. Theor. Biol. 180:27–37.

Taylor, P. D., G. Wild, and A. Gardner. 2007. Direct fitness or inclusive fitness: how shall we model kin selection. J. Evol. Biol. 20:301–309.

Wenseleers, T., and F. L. W. Ratnieks. 2006. Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. Am. Nat. 168:E163–E179.

Wenseleers, T., A. G. Hart, and F. L. W. Ratnieks. 2004a. When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. Am. Nat. 164:E154–E167.

Wenseleers, T., H. Helanterä, A. Hart, and F. L. W. Ratnieks. 2004b. Worker reproduction and policing in insect societies: an ESS analysis. J. Evol. Biol. 17:1035–1047.

Wenseleers, T., N. S. Badcock, K. Erven, A. Tofilski, F. S. Nascimento, A. G. Hart, T. A. Burke, M. E. Archer, and F. L. W. Ratnieks. 2005. A test of worker policing theory in an advanced eusocial wasp, *Vespula rufa*. Evolution 59:1306–1314.

West, S. A., and A. Buckling. 2003. Cooperation, virulence and siderophore production in bacterial parasites. Proc. R. Soc. Lond. B 270:37–44.

West, S. A., M. G. Murray, C. A. Machado, A. S. Griffin, and E. A. Herre. 2001. Testing Hamilton's rule with competition between relatives. Nature 409:510–513.

West, S. A., E. T. Kiers, E. L. Simms, and R. F. Denison. 2002a. Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proc. R. Soc. Lond. B 269:685–694.

West, S. A., E. T. Kiers, I. Pen, and R. F. Denison. 2002b. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? J. Evol. Biol. 15:830–837.

West, S. A., I. Pen, and A. S. Griffin. 2002c. Cooperation and competition between relatives. Science 296:72–75.

West, S. A., A. Gardner, D. M. Shuker, T. Reynolds, M. Burton-Chellow, E. M. Sykes, M. A. Guinnee, and A. S. Griffin. 2006. Cooperation and the scale of competition in humans. Curr. Biol. 16:1103–1106.

West, S. A., A. S. Griffin, and A. Gardner. 2007a. Evolutionary explanations for cooperation. Curr. Biol. 17:R661–R672.

———. 2007b. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20:415–432.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Yu D. W., J. Ridley, E. Jousselin, E. A. Herre, S. G. Compton, J. M. Cook, J. C. Moore, G. D. Weiblen 2004. Oviposition strategies, host coercion and the stable exploitation of figs by wasps. Proc. R. Soc. Lond. B 271:1185–1195.

Associate Editor: S. Gandon

Appendix

FRANK'S MODEL OF SELFISHNESS AND POLICING

Here, we analyze the dynamics and statics of Frank's (1995) model of the coevolution of selfishness and policing, deriving the results stated in the main text. In the context of this model, the average fitness of an individual is given by equation (1) of the main text. The direction of selection acting upon any trait is given by the marginal fitness with respect to that trait, that is, dw/dz and dw/da for selfishness and policing, respectively. These derivatives may be separated into partial derivatives, describing the effect of an individual's trait value, and the effect of the average trait value of its group, on this individual's fitness, using the chain rule methodology of Taylor and Frank (1996; see also Frank 1995, 1996b, 2003; Taylor 1996; Taylor et al. 2007). This gives $dw/dz = \partial w/\partial z + r\, \partial w/\partial z'$ and $dw/da = \partial w/\partial a + r\, \partial w/\partial a'$, where $r = dz'/dz = da'/da$ is the kin-selection coefficient of (whole-group) relatedness (Taylor and Frank 1996; Pepper 2000), or

$$\frac{dw}{dz} = (1-a)\left(\frac{(1-r)(1-(1-a)z)}{z} - r(1-ca)\right), \quad (A1)$$

and:

$$\frac{dw}{da} = rz(1-ca) - c(1-(1-a)z). \quad (A2)$$

First, we consider the evolution of selfishness in the absence of policing ($a = 0$). Evaluating the RHS of equation (A1) at $a = 0$ and $z \to +0$ gives $+\infty$, which is positive and hence selfishness is always favored in the absence of policing. The ESS level of selfishness z^* is determined by setting the right-hand side of equation (A1) to zero and solving for z; in the absence of policing ($a = 0$), this gives $z^* = 1 - r$. More generally, holding

the level of policing fixed at a, the ESS level of selfishness is given by $z^* = (1 - r)/[1 - a(1 - r(1 - c))]$. When there is full policing ($a^* = 1$) selfishness is neutral ($dw/dz = 0$), but in the limit of full policing ($a \to 1$), the ESS level of selfishness is $z^* \to (1 - r)/(1 - c)r$. This is stable as any reduction in selfishness ($z = [(1 - r)/(1 - c)r] - \delta$, where δ is a vanishingly small quantity), favors an increase in policing ($dw/da > 0$).

Next, we determine when policing can evolve, by evaluating the right-hand side of equation (A2) at $a = 0$ and $z = 1 - r$. This gives $dw/da = r(1 - r) - cr$, and hence the condition for policing to invade is $r < 1 - c$. Assuming that policing does evolve, we determine the evolutionary stability of full policing ($a = 1$) by evaluating the right-hand side of equation (A2) at $a = 1$ and $z = (1 - r)/(1 - c)r$, which obtains $dw/da = 1 - c - r$. Hence, full policing is evolutionarily stable ($a^* = 1$) if $r < 1 - c$. Setting the right-hand side of both equations (A1) and (A2) to zero, and simultaneously solving for a and z, yields no biologically feasible solutions; hence, an intermediate level of policing ($0 < a^* < 1$) is never evolutionarily stable. This means that when $r > 1 - c$, policing cannot evolve ($a^* = 0$), whereas when $r < 1 - c$, then policing can evolve, and the level of policing increases until full policing is established ($a^* = 1$). Policing favors an increase in selfishness: differentiating the ESS level of selfishness $z^* = (1 - r)/[1 - a(1 - r(1 - c))]$ with respect to a yields a positive slope ($dz^*/da = [(1 - r)(1 - (1 - c)r)]/[1 - a(1 - (1 - c)r)]^2 > 0$) when $r < 1/(1 - c)$ which is always met as $1/(1 - c) > 1$.

To investigate whether policing evolves due to direct or indirect fitness benefits, we must rearrange the marginal fitness for policing to be in the form of Hamilton's rule. Above, we expressed the marginal fitness for policing as $dw/da = dw/da + r(dw/da')$, where r is the whole-group relatedness (relatedness to average group member, including onself; Pepper 2000). We now rewrite in terms of others-only relatedness R (relatedness to average group member, not including self; Pepper 2000), which is the relatedness term that appears in a formal statement of Hamilton's rule. Substituting in the expression $r = 1/n + [(n - 1)/n]R$, we may rearrange equation (A2) into the form RB-C, where -C is the direct fitness effect and RB is the indirect fitness effect and their sum is the inclusive fitness effect (Hamilton 1963, 1964b). Hence, we have $B = [(n - 1)/n](dw/da')$ and $-C = dw/da + [(1/n)(dw/da')]$. For Frank's model, this obtains $B = [(n - 1)/n](z(1 - ac))$ and $C = c(1 - (1 - a)z) - [(1/n)(z(1 - ac))]$. We evaluate B and C at the ESS $a = a^* = 1$ and $z = z^* = (1 - r)/(1 - c)r$, which gives $B = [(n - 1)(n(1 - c) - 1)(1 - R)]/[(1 - c)n(1 + R(n - 1))]$ and

$C = c + 1/n - 1/(1 + R(n - 1))$. Social behaviors are classified according to the sign of B and C (Hamilton 1964a; Table 2): we

find that policing is altruistic ($B > 0$ and $C > 0$) if $R > [n(1 - c) - 1]/[(n - 1)(1 + cn)]$, and otherwise it is mutually beneficial ($B > 0$ and $C < 0$)—and this is Result 1 of the main text.

THE INDIVIDUAL AND GROUP COST OF SELFISHNESS

Here, we determine the evolutionarily stable levels of policing and selfishness for our modified model of policing. Note that the model presented in the section "The Group Cost of Selfishness" is a generalization of the model presented in "The Individual Cost of Policing." The results of both can be derived from a single model with Results 2 and 3 of the main text being given in the special case of $b = 1$. The average fitness of an individual is given by equation (4) of the main text. Marginal fitness for selfishness (z) and policing (a) are given by

$$\frac{dw}{dz} = \frac{(1 - r)(1 - a)}{z} - (1 - ab)(1 - a(1 - r)), \quad (A3)$$

and:

$$\frac{dw}{da} = brz - c. \quad (A4)$$

We first consider selfishness. Holding the level of policing fixed at a, and setting the RHS of equation (A3) equal to zero and solving for z, we obtain an ESS level of selfishness of $z^* = [(1 - a)(1 - r)]/[(1 - ab)(1 - a(1 - r))]$. If we assume there is no policing (set $a = 0$), as in Frank's model, this gives selfishness as $z^* = 1 - r$. The marginal fitness for policing at this point is given by evaluating equation (A4) at $z = 1 - r$, obtaining $dw/da = br(1 - r) - c$, and hence policing can invade only when $r(1 - r) > c/b$ (Results 2 and 6 of the main text). For full policing ($a = 1$) when $b = 1$, we find that selfishness is neutral ($dw/dz = 0$) at $a = 1$. In the limit of full policing ($a \to 1$) selfishness tends to $z^* \to (1 - r)/r$ (Result 2 and 3, main text). This ESS is stable as any reduction in selfishness ($z = (1 - r)/r - \delta$ where δ is a vanishingly small quantity) favors an increase in policing ($dw/da > 0$). For full policing ($a = 1$) when $b < 1$ we find that $dw/dz = -(1 - b)r$, so selfishness is disfavored and will decline to zero and this obtains $dw/da = -c$ which is always negative. Hence for $b < 1$, full policing is never stable (Result 4 of the main text). This means that when policing can invade from rarity ($r(1 - r) > c/b$), the ESS level of policing must be intermediate ($0 < a^* < 1$). To find this explicitly, we set the right-hand side of equations (A3) and (A4) to zero and simultaneously solve $a = a^*$ and $z = z^*$. This yields two results, only one of which is a biologically feasible ESS at $z^* = c/rb$ and

$$a^* = \frac{bc + (c - br)(1 - r) + \sqrt{4bc(1 - r)(br(1 - r) - c) + (bc + (c - br)(1 - r))^2}}{2bc(1 - r)}, \quad (A5)$$

(Result 4 and 8 of the main text). The maintenance and stability of this ESS (Results 6 of the main text) is discussed in the next

section. Using the approach outlined in the previous section, we may rearrange the right-hand side of equation (A4) into the form $RB-C$, where $B = bz(n-1)/n$ and $C = c - bz/n$. Evaluating this at the above ESS values of z^* & a^*, we find that $B > 0$ and $C > 0$, and hence policing is always altruistic in this variant model (Result 8 of the main text).

Finally, we examine the impact of b—the efficiency with which a policing behavior restores the group resources—upon the ESS level of policing.

Differentiating the RHS of equation (A5) with respect to b obtains:

$$\frac{da^*}{db} = \frac{bc - (c - br)(1 - r) - \sqrt{4bc(1 - r)(br(1 - r) - c) + (bc + (c - br)(1 - r))^2}}{2b^2\sqrt{4bc(1 - r)(br(1 - r) - c) + (bc + (c - br)(1 - r))^2}}, \quad (A6)$$

which is always positive, hence the ESS level of policing is stable at higher values as the efficiency of policing improves (see Fig. 2B).

DEMOGRAPHY AND POLICING—ILLUSTRATIVE OVERVIEW

Here, we consider the evolution of policing and selfishness in a demographic context, using the illustrative overview presented in Section "Demography and Policing—Illustrative Overview" of the main text where the average fitness of an individual is given by equation (5). The marginal fitness for selfishness (z) and policing (a) are given by

$$\frac{dw}{dz} = \frac{(1 - a)(1 - r) - (1 - ab)(1 - a(1 - r) - rs)z}{z(1 - ac - z(1 - ab))}, \quad (A7)$$

and:

$$\frac{dw}{da} = \frac{(1 - s)brz - c(1 - rs)}{1 + a(bz - c) - z}. \quad (A8)$$

First we consider the evolution of selfishness. The ESS level of selfishness (z^*) is found by setting the right-hand side of equation (A7) to zero and solving for z; in the absence of policing ($a = 0$), this gives $z^* = (1 - r)/(1 - rs)$. More generally, if the level of policing is held at a, the ESS level of selfishness is given by

$$z^* = \frac{(1 - r)(1 - a)}{(1 - ab)((1 - a)(1 - r) + (1 - s)r)}. \quad (A9)$$

We determine the evolutionary stability of policing by evaluating the right-hand side of equation (A8) at $a = 0$ and $z = (1 - r)/(1 - rs)$, which obtains an invasion condition for policing of $r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$. For full policing ($a = 1$), we find that $dw/dz = -[(1 - b)r(1 - s)]/[1 - c - (1 - b)z]$ which is zero for $b = 1$ and negative for all $b < 1$, so selfishness is neutral in the special case of fully efficient policing and is otherwise disfavored and is expected to decline to zero. Evaluating equation (A8) at

$z = 0$ and $a = 1$ yields $-[c(1 - rs)]/(1 - c)$, which is always negative and hence full policing ($a = 1$) is never stable. This means that when policing can invade from rarity ($r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$), the ESS level of policing must be intermediate ($0 < a^* < 1$). To find this ESS, we set both equations (A7) and (A8) equal to zero, and solve simultaneously for a and z at $a = a^*$ and $z = z^*$. This yields a single biologically feasible result of $z^* = c(1 - rs)/br(1 - s))$ and a rather complicated a^* (which can be recovered by replacing Queller's relatedness $(r - sr)/(1 - sr)$ for r in equation (A5)).

Policing can only invade from rarity when $c < br(1 - r)$ $[(1 - s)/(1 - rs)^2]$. To show policing may be maintained at higher costs than this, we evaluate the marginal fitness for policing (A8) at the ESS level of selfishness (A9) which yields

$$\frac{dw}{da} =$$
$$\frac{br(1 - a)(1 - r)(1 - s) - c(1 - ab)(1 - a(1 - r) - rs)(1 - rs)}{(1 - ab)(ca^2(1 - r) + r(1 - s) - a(c(1 - rs)))}. \quad (A10)$$

Evaluating equation (A10) for a cost marginally above that which allows policing to invade ($c = br(1 - r)[(1 - s)/(1 - rs)^2]$), we find policing may be favored ($dw/da > 0$) when $b > [r(1 - s)]/[1 - rs - a(1 - r)]$, which as $a \to 0$, simplifies to $b > \rho$ (where ρ is Queller's relatedness $(r - rs)/(1 - rs)$). As full policing is never favored (for $a = 1$, $dw/da < 0$ when $b < 1$), this means an ESS must exist between $0 < a < 1$ outside the region where policing can invade (Result 6 of the main text). We conduct a numerical investigation of the stability of this ESS using the methodology of Otto and Day (2007, pp. 237–242; results in Fig. 4) which confirms that when policing can invade, the ESS is always stable and that policing is maintained in regions where it cannot invade (particularly at lower relatedness and higher c/b ratios). When the ESS is stable, the eigenvalues are negative complex numbers indicating that policing and selfishness spiral in toward their ESS values.

To examine the effect of local competition s on the coevolution of selfishness and policing, the invasion condition $r(1 - r)[(1 - s)/(1 - rs)^2] > c/b$ is solved for r, to yield the maximum and minimum relatedness values that permit policing to invade from rarity. These bounds take the form $r_{max} = X + Y$ and $r_{min} = X - Y$ where

$$X = \frac{1 - s(1 - 2c/b)}{2(1 - s(1 - sc/b))}, \quad (A11)$$

and

$$Y = \frac{(1-s)\sqrt{1-4c/b}}{2\,(1-s(1-sc/b))}. \tag{A12}$$

To know whether local competition favors or disfavors the evolution of policing, we first examine how an increase in local competition affects the range of relatedness values over which policing may invade. This range evolve is given by $r_{max} - r_{min} = 2Y$. Differentiating Y with respect to s yields

$$\frac{dY}{ds} = -\frac{s(c/b)\,(1-\tfrac{1}{2}s)\,(1-s)\sqrt{1-4c/b}}{(1-s)(1-s(1-s(c/b)))^2}, \tag{A13}$$

which is negative when $c/b < \tfrac{1}{4}$;, which is a necessary (but not sufficient) condition for the evolution of policing, hence the range of relatedness values over which policing invades decreases with increasingly local competition ($d(r_{max} - r_{min})/ds < 0$; Result 9 of the main text).

If increasing local competition leads to policing invading at higher levels of relatedness, then dr_{max}/ds and dr_{min}/ds are positive. For these to be positive, both dX/ds must be positive and dX/ds must be greater than dY/ds. Differentiating X with respect to s yields

$$\frac{dX}{ds} = \frac{c/b(1-s+s^2(\tfrac{1}{2}-c/b))}{(1-s(1-sc/b))^2}, \tag{A14}$$

which is positive when $c/b > 0$. Setting the right-hand side of equation (A13) equal to the right-hand side of (A14) gives the condition that $dX/ds > dY/ds$ when $c/b > 0$. Hence as local competition increases, policing is favored to invade at higher levels of relatedness (Result 8 of the main text).

To investigate the effect of increased local competition on selfishness we differentiate z^* (A9) with respect to s, which yields

$$\frac{dz^*}{ds} = \frac{r(1-r)(1-a)}{(1-ab)(1-a(1-r)-rs)^2}. \tag{A15}$$

This is positive for all $0 < r < 1$, which is a necessary (but not sufficient) condition for the evolution of policing, which means selfishness increases as competition becomes more localized (Result 10 of the main text).

DEMOGRAPHY AND POLICING—ISLAND MODEL

We now consider the evolution of policing in an island model setting (Wright 1931; Taylor 1992), as described in the main text. We assume that the average fecundity F of a focal individual is given by equation (6), the average relative fecundity of all individuals in a group F' is given by equation (7), and the average relative fecundity of all individuals in the population $\bar{F}$ is given by equation (8) in the main text. The focal individual has KF offspring of which mKF disperse and $(1-m)KF$ remain in the natal group. Each dispersing offspring arrives in a new group with $(1-m)$

$\bar{F}nK$ natives and $m\bar{F}nK$ other immigrants. Therefore the chance that a particular dispersing offspring succeeds in obtaining one of the n breeding spaces is $n/[1+(1-m)\bar{F}nK + m\bar{F}nK] \approx 1/\bar{F}K$ assuming K is very large. Therefore, the expected fitness of the focal individual that is obtained through its dispersing offspring is $mFK/\bar{F}K = mF/\bar{F}$. Turning now to the $(1-m)FK$ nondispersing offspring of the focal individual, these find themselves competing for the n breeding spaces in the natal patch with $(1-m)F'K$ natives (including each other) and $m\bar{F}nK$ immigrants, therefore the total number of offspring is $(1-m)F'nK + m\bar{F}nK$. Hence the expected fitness of the focal individual that is obtained through its nondispersing offspring is $[(1-m)FnK]/[(1-m)F'nK + m\bar{F}nK] = [(1-m)F]/[(1-m)F' + m\bar{F}]$. Thus the total number of dispersing and nondispersing offspring of a focal individual (which is equal to an individual's fitness) is given by

$$w = m\frac{F}{\bar{F}} + \frac{F(1-m)}{(1-m)F' + m\bar{F}}. \tag{A16}$$

Using equation (A16), we substitute the fecundity values given in Section "Demography and Policing—Illustrative Overview" to calculate the marginal fitnesses for selfishness (z) and policing (a) where:

$$\frac{dw}{da} = \frac{brz(1-(1-m)^2) - c(1-r(1-m)^2)}{1 + a(bz-c) - z}, \tag{A17}$$

and

$$\frac{dw}{dz} = \frac{(1-a)(1-r) - (1-ab)(1-a(1-r)-r(1-m)^2)z}{z(1 + a(bz-c) - z)}. \tag{A18}$$

Note that equations (A7) and (A8) recover equations (A17) and (A18), respectively, if we make the substitution $s = (1-m)^2$, that is, in the context of the inelastic island model the scale of competition is equal to the probability that two randomly chosen group mates both fail to disperse (Gardner and West 2006). Relatedness r may also be expressed as a function of model parameters, by deriving a recursion for relatedness over successive generations and calculating its equilibrium value (Taylor 1992). In generation t, the expected relatedness of an individual to a randomly chosen member of its group (including itself), is $r_t = 1/n + ((n-1)/n)R_t$ where $1/n$ is the probability of choosing itself (in which case relatedness is 1) and $(n-1)/n$ is the probability of choosing another individual (in which case relatedness is R_t, the "others-only" relatedness). The only way for two different individuals from the same group to be related is if neither of them dispersed (which occurs with a probability of $(1-m)^2$), in which case their relatedness is equal to the whole-group relatedness of the previous generation (r_{t-1}). Hence, we may write the following recursion:

$$r_t = \frac{1}{n} + \frac{n-1}{n}(1-m)^2 r_{t-1}, \tag{A19}$$

which can be solved at equilibrium ($r_{t-1} = r_t = r$) to give an equilibrium relatedness of:

$$r = \frac{1}{n - (n-1)(1-m)^2},$$ (A20)

as determined by Taylor (1992). Substituting equation (A20) into equations (A17) and (A18) obtains:

$$\frac{dw}{da} = \frac{m(m-2)(cn(a+n(1-a)) + b(1-n+cna^2(n-1) + a(n-1-cn^2)))}{(1-ab)(1+2m(n-1) - m^2(n-1))(1+ca^2(n-1) - acn)}.$$ (A24)

$$\frac{dw}{dz} = \frac{(1-(1-m)^2)((1-a)(n-1) + (1-ab)(a(n-1) - n)z)}{z(1 + a(bz-c) - z)(1+(n-1)(1-(1-m)^2))},$$ (A21)

and:

$$\frac{dw}{da} = \frac{(1-(1-m)^2)(bz - cn)}{(1 + a(bz-c) - z)(1+(n-1)(1-(1-m)^2))}.$$ (A22)

We now determine the evolutionarily stable levels of policing and selfishness. Considering selfishness first, we find the ESS level of selfishness (z^*) by setting the right-hand side of equation (A21) to zero and solving for z whilst holding the level of policing at a; this gives

$$z^* = \frac{(1-a)(n-1)}{(1-ab)(a+n(1-a))}.$$ (A23)

In the absence of policing ($a = 0$), this is $z^* = (n-1)/n$. Evaluating equation (A22) at $z^* = (n-1)/n$, gives an invasion condition for the evolution of policing of $(n-1)/n^2 > c/b$ (Result 11 of the main text). To find the stable level of policing, we set both equations (A21) and (A22) equal to zero, and solve simultaneously for a and z at $a = a^*$ and $z = z^*$. This yields one

biologically feasible solution of $z^* = (c/b)n$ and a complicated a^*, which can be recovered by substituting $\rho = 1/n = r$ in equation (A5) (Result 12 of the main text). To show that once present, policing may be maintained in larger groups we evaluate the marginal fitness for policing (A22) at the ESS level of selfishness (A23), which yields

Evaluating equation (A24) at $c = (b(n-1))/n^2$, (policing can only invade when $c < (b(n-1))/n^2$), we find policing is still favored ($dw/da > 0$) when $b > 1/(a + n(1-a))$ (which is equivalent to the open model result as in the island model, Queller's relatedness $\rho = 1/n$). As full policing is not stable (at $a = 1$, $dw/da < 0$), an intermediate ESS ($0 < a < 1$) must occur outside the region where policing can invade (Result 12 of the main text). We conduct a stability analysis for the Island model ESS using the methodology of Otto and Day (2007, pp. 237–242). The results can be obtained from Figure 4 by substituting s for $(1-m)^2$ and r for $1/[n - (n-1)s]$.

To investigate how changing migration rates affect both relatedness and the intensity of local competition, we first differentiate relatedness ($r = 1/(n - (n-1)(1-m)^2)$) with respect to m, which yields $dr/dm = -[2(1-m)(n-1)]/[1 + (1-(1-m)^2)(n-1)]^2$. This is always negative, so increasing the migration rate decreases relatedness. Differentiating the scale of competition term, $s = (1-m)^2$, in terms of m gives $ds/dm = -2(1-m)$ which is also always negative meaning that increasing migration decreases local competition. Both these effects must cancel resulting in the migration rate m having no net effect on the evolution of policing as a^* and z^* and ρ are not functions of m, so $da^*/dm = dz^*/dm = d\rho/dm = 0$ (Result 11, main text).