

Room at the top: biological lock-in of social advantage

David Hugh-Jones & Abdel Abdellaoui

2020-07-27

1 Introduction

Charles Murray (1995) warned of “a merging of the cognitive elite with the affluent”. On the opposite side of the political spectrum, Karl Marx (1844) wrote “I am ugly, but I can buy the most beautiful woman.... the effect of ugliness, its repelling power, is destroyed by money.” These quotations suggest that social advantages, such as wealth, caste or status, may be transformed into biological advantages in the next generation, via assortative mating between socially and genetically advantaged people. We call this process biological lock-in.

Figures 1 and 2 illustrate the idea using data for spouse pairs from UK Biobank. Figures 1 plots one partner’s mean polygenic score for educational attainment (PSEA) against a measure of the other partner’s actual educational attainment: possession of a university degree. University graduates had spouses with higher PSEA.¹ Figure 2 plots one partner’s PSEA against another measure of social status: income.

These figures do not prove that biological lock-in is taking place: since an individual’s own PSEA correlates with both their educational attainment, and their income, both figures could be a result of partner selection on a purely genetic basis. In this paper, we test the theory more rigorously, using environmental shocks to social status that are unlikely to be correlated with own genetics. First, we develop a simple theory of biological lock-in, to illustrate how its effects vary with social structure.

2 Theory

The population is size 1. There is a single binary genetic trait and a single social trait. Individuals i thus have $G_i \in \{0, 1\}$ and $S_i \in \{0, 1\}$; write $\tau_i = (G_i, S_i)$ to fully describe i ’s type. We will call individuals with $\tau_i = (1, 1), (1, 0), (0, 1)$ and $(0, 0)$, b-types, g-types, s-types and n-types respectively. The proportion of the population with $G_i = 1$ is denoted by γ , and the population proportion with $S_i = 1$ is denoted by σ . In generation 1, G_i and S_i are assumed independent, so e.g. the proportion of b-types is $\gamma\sigma$.

Higher values are assumed to be more attractive in the marriage market. Indeed, individuals are divided into two classes, H and L , for the purpose of marital matching. Within H and L , all individuals form spouse pairs at random. Class H is of size θ : we assume $\gamma\sigma < \theta < \min\{\gamma, \sigma\}$. We also assume γ and σ are not too different:

$$\text{Either } \sigma < \gamma < (1 + \sigma)/2 \text{ or } \gamma < \sigma < (1 + \gamma)/2. \quad (1)$$

Married pairs have two children and each randomly inherits one parent’s genetic trait and one parent’s social trait. As a result, σ and γ stay constant over generations.

¹To minimize concerns about genetic stratification, i.e. correlations between genetics and non-genetic forms of inherited advantage, PSEA is residualized by the first 100 principal components of UK Biobank array data.

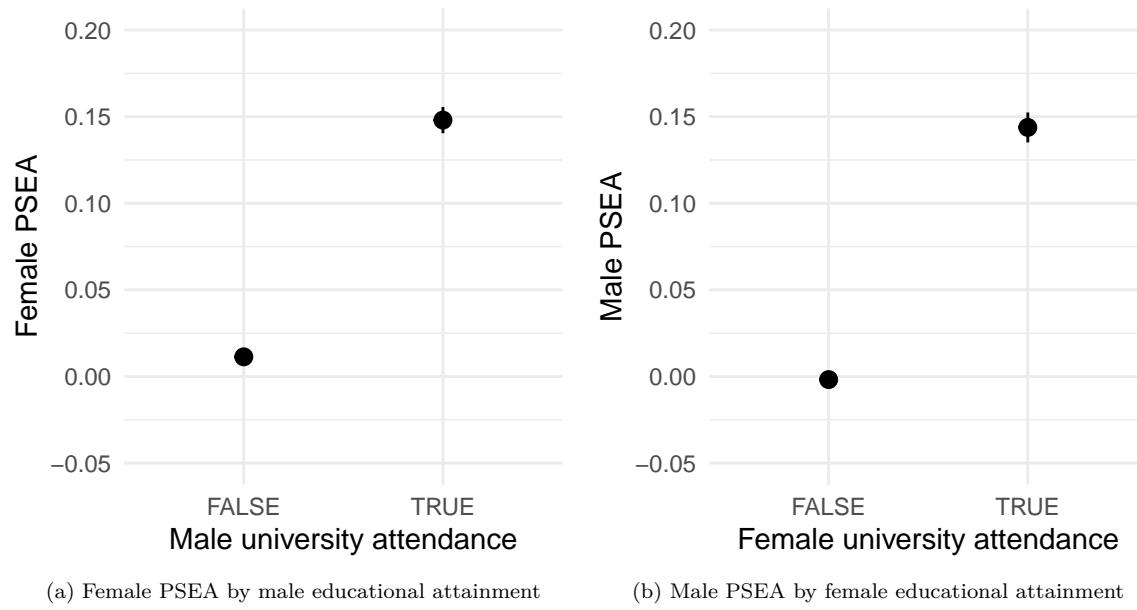


Figure 1: Social and genetic advantage among spouse pairs in UK Biobank

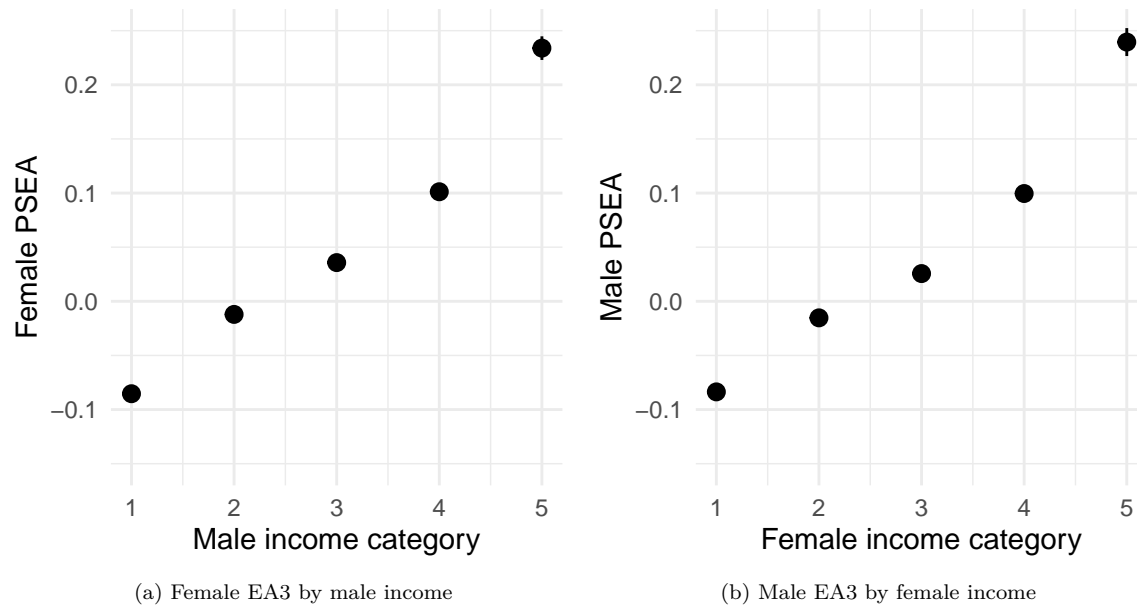


Figure 2: Social and genetic advantage among spouse pairs in UK Biobank

B-types with $\tau_i = (1, 1)$ are always in H . N-types with $\tau_i = (0, 0)$ are always in L . Where the other groups fit depends on the nature of the society. Of the $\phi \equiv \theta - \gamma\sigma$ non-b-types in H , a proportion k are s-types and the remaining $1 - k$ are g-types.

The value of k gives the level of social stratification in the society. To build intuition, we start with two extreme examples. Suppose first that $k = 1$ and that $\theta = \sigma$. This is a caste society where marital matching is entirely driven by social status. All individuals in H (L) have $S_i = 1$ ($S_i = 0$), and so do their children. On the other hand, the proportion of individuals with $G_i = 1$ stays at γ within both groups. Thus, over generations, the distributions of G_i and S_i remain independent in the population.

Suppose next that $k = 0$ and $\theta = \gamma$. This society is socially egalitarian: only genetics matter for marital matching. All individuals in H (L) have $G_i = 1$ ($G_i = 0$), and so do their children. The proportion of individuals with $S_i = 1$ stays at σ within both groups; hence, again, G_i and S_i remain independent over generations.

In between these cases, G_i and S_i will become positively correlated. We examine the correlation in the second generation.

Proposition 1. *The second-generation correlation between G_i and S_i is concave in k . If θ is close enough to $\min\{\sigma, \gamma\}$, then the correlation has a maximum at $k \in (0, 1)$.*

The logic behind Proposition 1 is simple. As we have seen, when spouse matching is entirely on the basis of social status, or entirely on the basis of genetics, no correlation arises between the two. However, when spouses are selected on the basis of both genetics and social status, then some s-types, who only have high social status, marry some g-types, who only have valuable genetic capital. As a result, some of their children inherit both (or neither), and this increases the correlation between genetics and social status.

Figure 3 plots the second generation correlation coefficient of S_i and G_i against k , for $\gamma = \sigma = \theta = 0.5$.

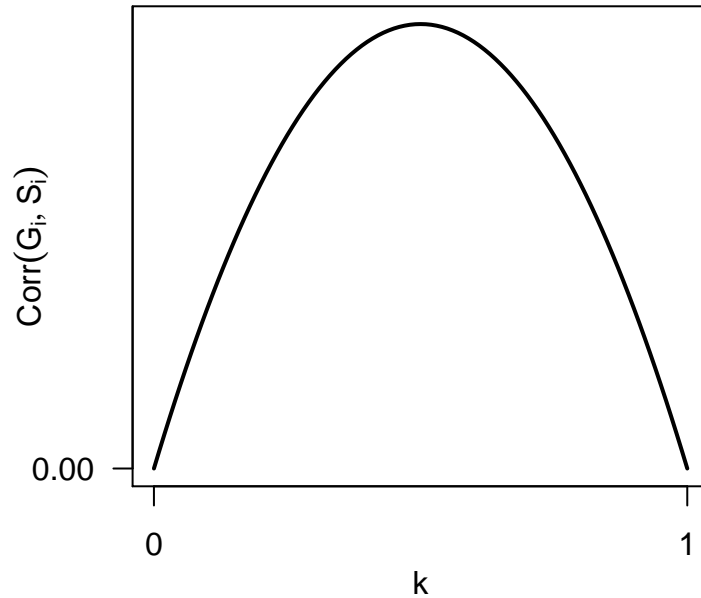


Figure 3: Social stratification and biological lock-in: theory

3 Data

As mentioned above, simple correlations between one partner's social status and the other partner's genetics do not prove that biological lock-in is taking place, because one's social status correlates with one's own genetics. To demonstrate biological lock-in, we therefore need a source of social advantage which is exogenous to genetics. One possibility is birth order. It is well known that earlier-born children receive more parental care and have better life outcomes. (XXX is it? Go check.) On the other hand, early- and late-born full siblings have the same ex ante expected genetic endowment.² We can therefore use birth order as an instrument for social advantage.

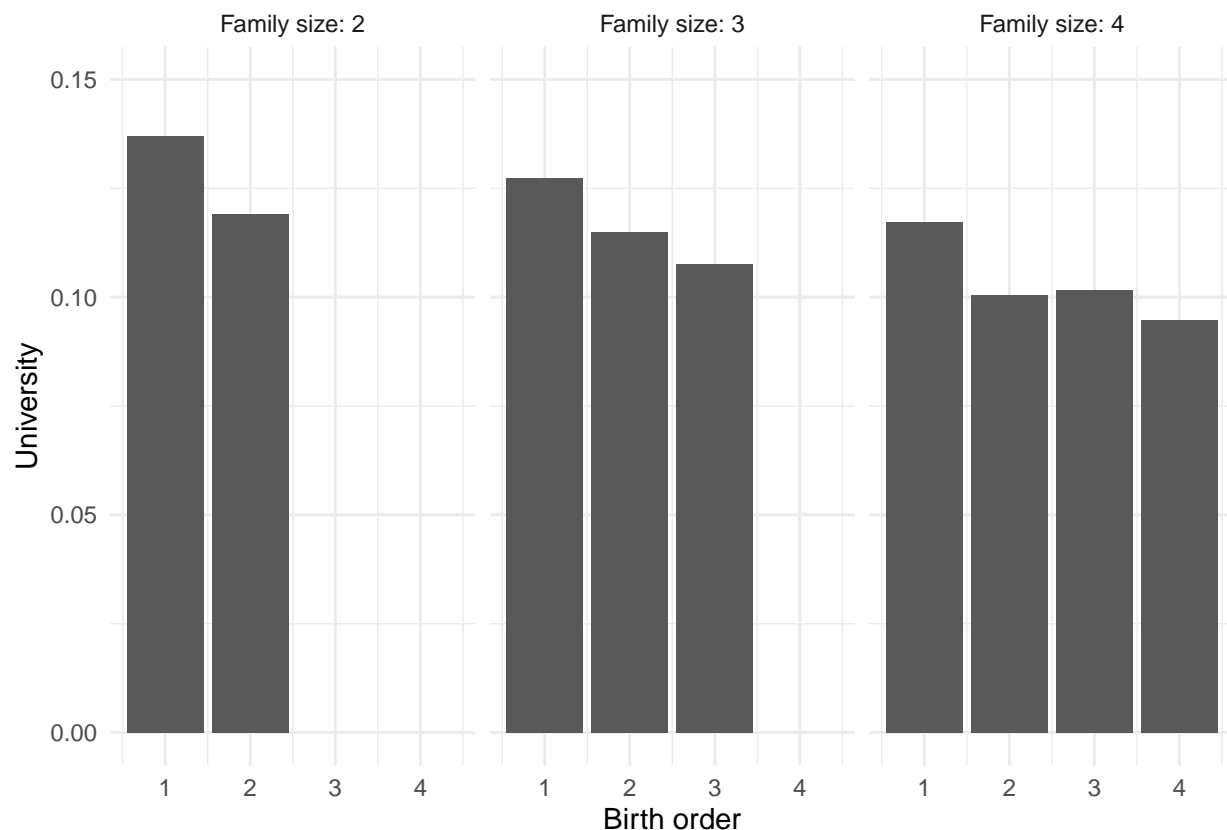
We use data from UK Biobank, a study of about 500,000 individuals.

TODO: describe N for birth order, describe PSEA calculation.

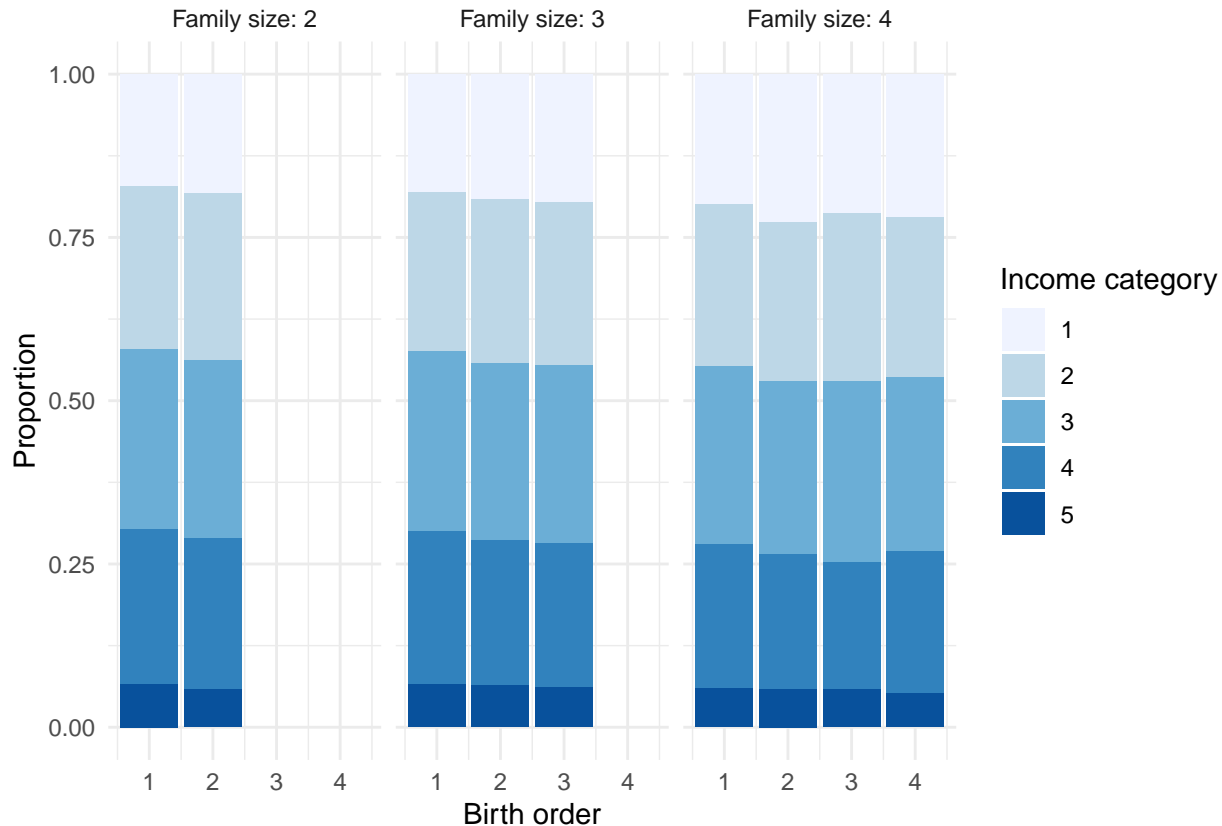
4 Results

TODO: reduced form

```
## `summarise()` regrouping output by 'Birth order' (override with ` .groups ` argument)
```



²This might not be the case, if parents' choice of whether to have more children is endogenous to the genetic endowment of their earlier children. We will check for this below.



4.1 Regressions

We run regressions instrumenting one spouse’s educational attainment with their birth order. We control for overall family (i.e. sibling group) size. Without this, then birth order would correlate with family size, which itself may correlate with genetics, since parents of different family sizes might have different genetic endowments. Indeed, respondents with more siblings overall have substantially lower PSEA. We also control for subject’s own PSEA. Lastly, we allow the effect of birth order to vary depending on total family size.

Table 1 shows instrumental variables regressions using 2 stage least squares. We use only subjects with between 2 and 8 siblings. Column 1 pools men and women. Columns 2 and 3 run separate regressions by sex. The effect is positive for both male and female subjects, but twice as strong, and significant, for males.

Although all children of the same parents have identical expected values of polygenic scores, it is conceivable that parents might select family size on the basis of genetics. For example, if the first child had a phenotype reflecting a high (or low) polygenic score, then that might affect the parents’ decision to have a second child. To check this we regress birth order directly on PSEA. As before, we control for family size and allow the effect of birth order to vary depending on family size. Birth order variables are not jointly significant (anova p 0.178) and their effect sizes are of the order of 0.01 of a standard deviation of PSEA.

Table 1: Effects of educational attainment on spouse's EA3

	All spouses' EA3	Female spouses' EA3	Male spouses' EA3
University	0.440 ** (0.161)	0.351 (0.187)	0.610 * (0.295)
EA3	0.033 *** (0.005)	0.035 *** (0.007)	0.029 ** (0.009)
Family size dummies	Y	Y	Y
R2	-0.007	-0.002	-0.019
N	340061	170663	169398
Sargan test (p)	0.467	0.407	0.844

*** p < 0.001; ** p < 0.01; * p < 0.05. Educational attainment instrumented by birth order.

5 Appendix

5.1 Proof of Proposition 1

Write b_t, g_t, s_t, n_t for the population proportion of each type in generation $t = 1, 2, \dots$. Sometimes we drop the subscript. Since G_i and S_i are independent in generation 1,

$$\begin{aligned}
b_1 &= \gamma\sigma; \\
g_1 &= \gamma(1 - \sigma); \\
s_1 &= (1 - \gamma)\sigma; \\
n_1 &= (1 - \gamma)(1 - \sigma).
\end{aligned}$$

Within either group $J \in \{H, L\}$, write the proportion of types in generation t as b_J, g_J, s_J, n_J for the respective types.

Dividing children up by their parents' types, we can calculate the proportion with each of the four types:

	b_J	g_J	s_J	n_J
b_J	$1; 0; 0; 0$	$\frac{1}{2}; \frac{1}{2}; 0; 0$	$\frac{1}{2}; 0; \frac{1}{2}; 0$	$\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}$
g_J	$\frac{1}{2}; \frac{1}{2}; 0; 0$	$0; 1; 0; 0$	$\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}$	$0; \frac{1}{2}; 0; \frac{1}{2}$
s_J	$\frac{1}{2}; 0; \frac{1}{2}; 0$	$\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}$	$0; 0; 1; 0$	$0; 0; \frac{1}{2}; \frac{1}{2}$
n_J	$\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}$	$0; \frac{1}{2}; 0; \frac{1}{2}$	$0; 0; \frac{1}{2}; \frac{1}{2}$	$0; 0; 0; 0$

Table 2: Proportions of children born from different types of parents (b-type; g-type; s-type; n-type)

Summing these in each group, and using that $n_H = 0$ and $b_L = 0$, gives the proportion of children of each

type in each group:

$$\begin{aligned}b_{J,t+1} &= b_J + \rho_J \\g_{J,t+1} &= g_J - \rho_J \\s_{J,t+1} &= s_J - \rho_J \\n_{J,t+1} &= n_J + \rho_J\end{aligned}$$

where

$$\rho_J \equiv \frac{1}{2}g_J s_J.$$

In words, when a group contains both g-types and s-types, some of them marry each other and produce some b- and n-type offspring: this increases the correlation between G_i and S_i .

Summing over the groups, we can write the new generation's proportion of types in the whole population as

$$\begin{aligned}b_{t+1} &= b_t + \rho \\g_{t+1} &= g_t - \rho \\s_{t+1} &= s_t - \rho \\n_{t+1} &= n_t + \rho\end{aligned}$$

where

$$\rho \equiv \theta \rho_H + (1 - \theta) \rho_L.$$

We can now calculate the covariance between G and S in the second generation. The covariance is just $E(G_i S_i) - E(G_i)E(S_i)$; in general, at generation t , this is $b_t - \gamma\sigma$. For $t = 2$ this is just equal to ρ .

In the first generation, we can calculate:

$$\begin{aligned}b_{H,2} &= \gamma\sigma/\theta; \\g_{H,2} &= (1 - k)\phi/\theta \\s_{H,2} &= k\phi/\theta; \\n_{H,2} &= 0;\end{aligned}$$

and

$$\begin{aligned}b_{L,2} &= 0; \\g_{L,2} &= \frac{\gamma(1 - \sigma) - \phi(1 - k)}{1 - \theta} \\s_{L,2} &= \frac{(1 - \gamma)\sigma - \phi k}{1 - \theta} \\n_{L,2} &= \frac{(1 - \gamma)(1 - \sigma)}{1 - \theta};\end{aligned}$$

and from these

$$\begin{aligned}\rho_H &= \frac{1}{2} \frac{\phi^2 k(1 - k)}{\theta^2}; \\ \rho_L &= \frac{1}{2} \frac{[\gamma(1 - \sigma) - \phi(1 - k)][(1 - \gamma)\sigma - \phi k]}{(1 - \theta)^2}.\end{aligned}$$

Hence:

$$\rho = \frac{1}{2} \frac{\phi^2 k(1-k)}{\theta} + \frac{1}{2} \frac{[\gamma(1-\sigma) - \phi(1-k)][(1-\gamma)\sigma - \phi k]}{(1-\theta)}.$$

Note that when $k = 1$ and $\theta = \sigma$, ρ evaluates to 0. Similarly, when $k = 0$ and $\theta = \gamma$, ρ evaluates to 0. This confirms the claims in the main text about the two extreme example societies.

ρ is a quadratic in k . Differentiating ρ by k gives

$$\begin{aligned} \frac{d\rho}{dk} &= \frac{1}{2\theta} \phi^2(1-2k) + \frac{1}{2(1-\theta)} [\phi^2(1-2k) + \phi(\sigma - \gamma)] \\ &= \frac{1}{2\theta(1-\theta)} \phi^2(1-2k) + \frac{1}{2(1-\theta)} \phi(\sigma - \gamma) \\ &= \frac{1}{2(1-\theta)} \left[\frac{\phi^2}{\theta} (1-2k) + \phi(\sigma - \gamma) \right]. \end{aligned} \tag{2}$$

This is decreasing in k , proving that ρ is concave.

We find the maximum for ρ by solving the first order condition. In (2), setting the term in square brackets to 0 gives:

$$k = \frac{1}{2} + \frac{1}{2} \frac{\theta}{\theta - \gamma\sigma} (\sigma - \gamma). \tag{3}$$

We need to show this is strictly between 0 and 1 as θ approaches σ or γ .

First assume $\sigma < \gamma$. As $\theta \rightarrow \sigma$, the expression (3) approaches $k = \frac{1}{2} + \frac{1}{2} \frac{(\sigma - \gamma)}{(1 - \gamma)}$. The requirement for this to be interior is that $|\sigma - \gamma| < |1 - \gamma|$. For $\sigma < \gamma$ this is equivalent to $\gamma < (1 + \sigma)/2$, which holds by condition (1).

Next assume $\gamma < \sigma$. As $\theta \rightarrow \gamma$, (3) approaches $k = \frac{1}{2} + \frac{1}{2} \frac{(\sigma - \gamma)}{(1 - \sigma)}$. This is interior when $|\sigma - \gamma| < |1 - \sigma|$, equivalently when $\sigma < (1 + \gamma)/2$, which again holds by (1).