NATURAL SCIENCES TRIPOS **Part IA**

Friday 9th June 2023 1:30 to 4:30 pm

MATHEMATICAL BIOLOGY

*You have **three hours** to answer this paper (plus any pre-agreed individual adjustments).*

You must answer eight questions, including at least one question from each of Sections A to E. Indicate your question selection on the form to the right.

All questions carry equal weight. Indicative proportions of marks are given for each question part.

You may use an approved calculator and the Mathematical Biology Formula Booklet.

*Begin each answer on a **separate** blank page.*

*Answers should be **hand-written,** either electronically on a tablet or on paper and then uploaded to Inspera. Make sure that your handwriting and any diagrams or equations are clear and legible in your submitted files.*

Where relevant, formulae, equations, diagrams, graphs and sketches should be drawn by hand, not prepared using a computer. Do not copy and paste any figures from other documents.

For full credit, calculation steps must be shown where necessary to explain your answers.

*This examination **closed book** so access to the internet and use of mobile devices is prohibited.*

After you have completed the assessment:

*Your answers should be submitted **using either Inspera scan (see below) or by digital upload using the Inspera Upload Tool. Both options are available at the end of the paper.***

Note:

- The single Inspera Scan question code (at the end of the paper) **MUST** be noted during the exam.

- Please ensure that each Inspera Scan page is labelled appropriately, including your Blind Grade Number.
- Inspera Scan pages **MUST** be placed in the Inspera Scan folder at the completion of the exam, and left on your desk.
- Inspera Scan pages will be scanned by administrators after the completion of the exam.
- If you are using the Inspera Upload Tool, answers should be uploaded as a single document.

Stationery requirements:

- Rough work pad
- Inspera Scan pages
- Inspera Scan Folder

SECTION A

A1

(a) The random variable y has the following probability density function:

$$f(y) = \begin{cases} k(y^2 + 2y + 2) & \text{for } 1 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

where *k* is a positive constant.

(i) Calculate the value of *k.*

[~15% marks]

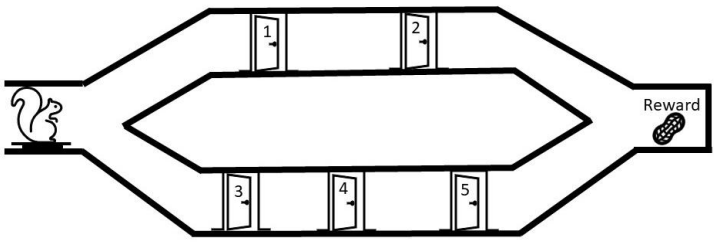
(ii) Derive the expression of the cumulative distribution function of *y.*

[~15% marks]

(iii) Evaluate $P(y \leq 2)$.

[~10% marks]

(b) A gardener sets up a challenge for visiting squirrels. When a squirrel stands on the starting platform, each of 5 doors in two parallel tubes are randomly set as 'open' with probability p_i and 'shut' with probability $q_i = 1 - p_i$ where $i = 1 \ldots 5$ is the door number as shown in the diagram:



(i) Write down an expression for the probability that, in any given trial, the squirrel will be able to reach the reward in the end chamber (assume that if the squirrel reaches a closed door in one tunnel it will try the other tunnel).

[~15% marks]

The experiment is set up so that the probability of reaching the reward is 0.15. After each trial, the squirrel exits the contraption before returning for another go (resetting it each time).

(ii) What is the expected number of attempts required for the squirrel to get the reward? Explain your answer.

[~15% marks]

(iii) What is the probability that it takes at least three attempts for the squirrel to get the reward?

[~10% marks]

(iv) The squirrel manages to collect and bury 50 nuts over an area of $100m^2$. Another squirrel comes to the same patch of land hoping to steal some nuts. If the thief searches an area of $4m^2$, what is the probability that it finds exactly 5 nuts? What assumptions did you make?

[~20% marks]

A2

(a) A microfluidic mixing device has four inputs and four outputs, connected as follows:

Flow into input 1 goes to outputs 1 and 2.
Flow into input 2 goes to outputs 1 and 3.
Flow into input 3 goes to outputs 2 and 4.
Flow into input 4 goes to outputs 2, 3 and 4.

In each case, input flow is divided evenly among the corresponding outputs.

(i) These connections can be summarised by a matrix **M**, where M_{ij} is the proportion of flow from input j which goes to output i. Write down this matrix for the device specified here.

[~16% marks]

(ii) Let **x** be a vector whose i^{th} element is the flow into input i, and let **y** be a vector whose i^{th} element is the total flow from output i. Write down a matrix-vector equation relating **M**, **x** and **y**, and solve it to find the input flows in the case where the following output flows are measured from the device:

output 1: 3.0 µl/s

output 2: 10.0 µl/s

output 3: 1.0 µl/s

output 4: 7.0 µl/s

[~24% marks]

(b) Calculate a complete set of eigenvalues and eigenvectors for the following matrix:

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$$

[~60% marks]

SECTION B

B3

An oncologist wants to test whether a given mutation is associated with the development of prostate cancer. She types a set of patients with the disease and a set of control individuals that do not show any symptoms, scoring individuals as either carriers of the mutation or not (i.e. with the ancestral version of the allele). She gets the following results:

	mutation	**ancestral**
cancer	736	1245
healthy	874	1345

(a) Test whether carrying the mutation is associated with the disease. Specify which test is appropriate and show your working.

[~15% marks]

(b) Write a sentence to report your results in a paper.

[~10% marks]

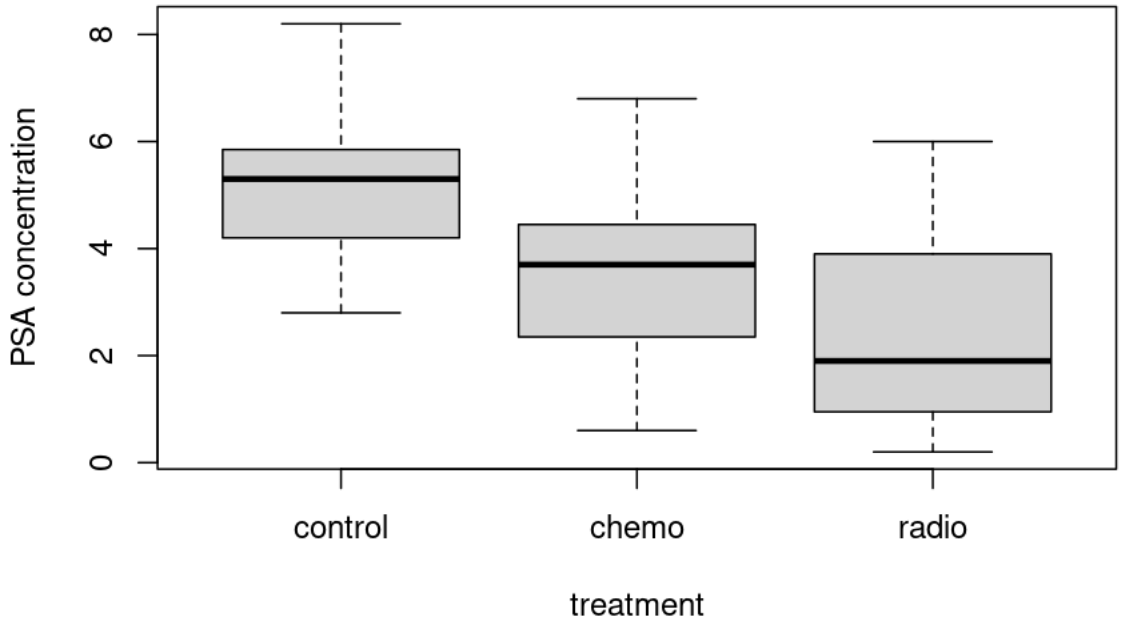
(c) From a previously published study, she knows that 40% of individuals carry the mutation in another population. Does the frequency of carriers of the mutation in the current study differ from that reported in the literature? Specify which test is appropriate and show your workings.

[~15% marks]

(d) Write a sentence to report your results in a paper.

[~10% marks]

Our oncologist would also like to test the impact of different treatments used to reduce the likelihood of cancer reoccurrence after surgery. She compares control individuals (who were provided with a placebo treatment) with patients who received chemo- and radiotherapy. After 6 months, the levels of prostate-specific antigen (PSA) in the blood were measured. Individuals with higher levels of PSA have a greater risk of the cancer reoccurring. She obtains the following results:



(e) She performs a one-way ANOVA on the PSA concentration to compare the treatments, and she obtains the table:

	Df	Sum Sq	Mean Sq	F value
Treatment	**??**	64.454	**??**	**??**
Residuals	42	**??**	**??**	
Total	44	176.226	34.888	

Complete the cells with **??** (Df, Sum Sq, Mean Sq, F), and test the significance. What do you conclude?

[~20% marks]

(f) She then performs a post-hoc test and gets the following output:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = psa_lm)

$treatment
                     diff         lwr         upr
chemo-control      -1.940      -3.387      -0.493
radio-control      -2.873      -4.321      -1.426
radio-chemo        -0.933      -2.38        0.514
```

What do you conclude?

[~15% marks]

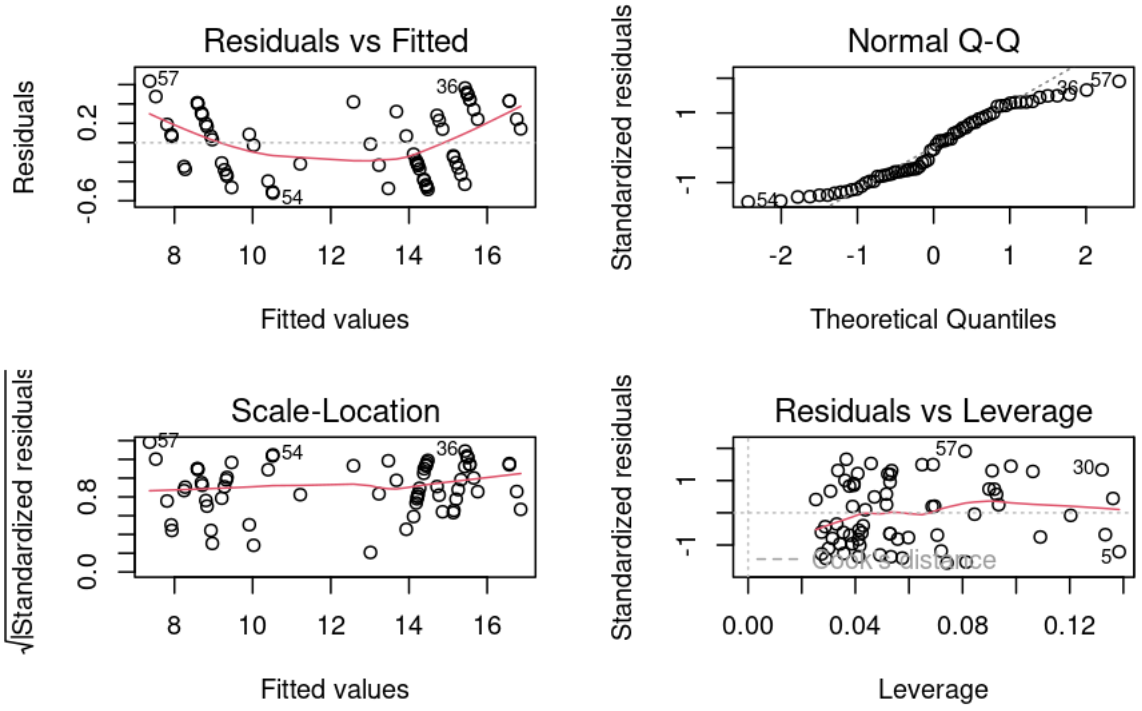
(g) Write a sentence to report the results from **(e)** and **(f)** in a paper.

[~15% marks]

B4

A geneticist wants to test whether genetic heterozygosity affects reproductive success. He collects data from pheasants, counting the number of hatched chicks as a measure of success. Besides measuring heterozygosity, he also collects data on laying date (given as the number of days since the 1st of March), and notes whether the hens were wild-bred or captive-bred. There were 67 nests in total.

He begins the statistical analysis with a linear model, but looking at the diagnostic plots shown below, he decides that the model is not appropriate.



(a) Is he correct? Are the diagnostic plots concerning?

[~20% marks]

He decides instead to fit a Poisson GLM with a log link function and gets the following output:

```
Call:
glm(formula = n_clutch ~ het + breeding + n_days, family = "poisson", data =
pheasants)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-0.166357  -0.070218  -0.009454   0.073792   0.177674

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.888261   0.185415  10.184  < 2e-16 ***
het          0.840696   0.389046   2.161   0.0307 *
breedingwild 0.492778   0.076792   6.417 1.39e-10 ***
n_days       0.002181   0.005567   0.392   0.6953
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 51.02136  on ??  degrees of freedom
Residual deviance:  0.42248  on ??  degrees of freedom
AIC: 299.48
```

(b) Calculate the degrees of freedom of the null and residual deviance.

[~10% marks]

(c) What is the expected number of chicks in a nest for a hen that was wild-bred, had a heterozygosity of 0.4, and laid eggs 25 days after the 1st of March? And for a captive-bred hen with the same characteristics?

[~10% marks]

(d) Dropping each of the predictors from the full model, he gets:

Single term deletions

Model:
n_clutch ~ het + breeding + n_days

	Df	Deviance	AIC	LRT
<none>		0.422	299.48	
het	1	5.089	302.14	**??**
breeding	1	44.079	341.13	**??**
n_days	1	0.576	297.63	**??**

Compute the appropriate LRT values for each predictor and test their significance. What would be the next step?

[~20% marks]

(e) He decides to drop n_days, the number of days since 1st of March (this may or may not be correct based on c; do not trust his decision to inform your answer to question c):

```
Single term deletions

Model:
n_clutch ~ het + breeding
          Df    Deviance    AIC      LRT
<none>           0.576    297.63
het        1     5.362    300.41     ??
breeding   1    45.748    340.80     ??
```

Compute the appropriate LRT values for each predictor and test their significance. What would be the next step?

[~15% marks]

(f) Write a sentence to summarise the results for a paper.

[~10% marks]

(g) He then decides to take an information criterion approach to compare models and obtains:

Fixed term is "(Intercept)"
Global model call: glm(formula = n_clutch ~ het + breeding + n_days, family = "poisson", data = pheasants)

Model selection table

	(Int)	brd	het	n_dys	df	logLik	AICc	delta	weight
4	1.938	+	0.8505		3	-145.815	298.0	0.00	0.579
8	1.888	+	0.8407	0.002181	4	-145.738	300.1	2.11	0.202
2	2.197	+			2	-148.207	300.6	2.59	0.158
6	2.127	+		0.002930	3	-148.071	302.5	4.51	0.061
3	2.250		0.8931		2	-168.401	341.0	42.98	0.000
7	2.078		0.8596	0.007214	3	-167.566	341.5	43.50	0.000
1	2.524				1	-171.037	344.1	46.13	0.000
5	2.321			0.008041	2	-170.010	344.2	46.20	0.000

Models ranked by AICc(x)

If he applies a threshold of 4 to deltaAIC, what models should be included in the most supported set? Explain how a weight of 0.579 should be interpreted.

[~15% marks]

SECTION C

C5

(a) Calculate the derivative of

$$Y(t) = \log(p + e^{-qt})$$

where p and q are positive constants. (Note: here log refers to the natural logarithm, i.e., base e)

[~10% marks]

A certain species of fungus causes disease in wheat. This fungus persists in agricultural soil via hardened masses of hyphae called sclerotia. Mycelial growth from sclerotia can lead to infection of plants. Within a growing season, the fungus increases in density on the roots of infected plants. However, no new sclerotia are formed when the crop is present. It is harvesting the wheat crop that causes new sclerotia to be released to the soil. Secondary infection, i.e. the direct spread of the fungus from one wheat plant to another within a growing season, is so rare that it can reasonably be ignored.

The density of sclerotia in a field during a crop growing season, $X(t)$, can be modelled as

$$\frac{dX}{dt} = -\gamma X,$$

in which t is the time since crop planting and γ is a positive parameter. The initial density of sclerotia is $X(0) = X_0$.

(b) Briefly explain the model, giving a possible biological interpretation of γ.

[~5% marks]

(c) Solve the model to find an expression for $X(t)$.

[~10% marks]

A model of the number of infected wheat plants, $I(t)$, at time t during the growing season is:

$$\frac{dI}{dt} = \frac{\beta X}{\alpha + X}(N - I),$$

in which N is the (constant) number of wheat plants, X is the (time-dependent) density of sclerotia, and α and β are positive parameters. No wheat plants are infected initially.

(d) Explain the model, giving a possible biological interpretation of α and β.

[~20% marks]

(e) Substitute the expression for $X(t)$ from question (c) into the model and solve the differential equation for $I(t)$ to show that:

$$I(t) = N\left(1 - \left(\frac{\alpha + X_0}{\alpha + X_0 e^{-\gamma t}}\right)^{-\frac{\beta}{\gamma}}\right).$$

*(Hint. You might find your answer to **(a)** to be helpful)*

[~30% marks]

(f) Verify that the expression given in ***(e)*** is indeed a solution of the model.

[~25% marks]

C6

The differential equation

$$\frac{dY}{dt} = F(Y)$$

has an equilibrium at $Y = \bar{Y}$.

(a) By considering the dynamics of a perturbation, $h(t)$, with $Y(t) = \bar{Y} + h(t)$, and $h(0) = h_0$, where h_0 is very small, show how the stability of the equilibrium depends on the sign of dF/dY evaluated at $Y = \bar{Y}$.

[~25% marks]

A model of the size, $Y(t)$, of a population at time t is

$$\frac{dY}{dt} = G(Y) = \beta Y\left(1 - \left(\frac{Y}{\kappa}\right)^2\right),$$

in which β and κ are positive parameters.

(b) Using appropriate technical vocabulary, explain the model. What are the meanings of the parameters β and κ?

[~10% marks]

(c) Sketch $G(Y)$ as a function of Y.

[~10% marks]

(d) Calculate the population size at which the population grows most rapidly.

[~15% marks]

An extension of the above model accounts for predation at net rate $H(Y)$, with:

$$\frac{dY}{dt} = G(Y) - H(Y) = \beta Y\left(1 - \left(\frac{Y}{\kappa}\right)^2\right) - \alpha Y,$$

in which α is a positive parameter.

For the remainder of this question, you can assume $\alpha < \beta$, although your answers should clearly state whenever you have used this information.

(e) Find every biologically meaningful equilibrium value of the extended version of the model, and use the result obtained in ***(a)*** to assess their stability.

(*Hint: "Biologically meaningful" means you need only consider* $Y \geq 0.$)

[~15% marks]

(f) By sketching $G(Y)$ and $H(Y)$ on the same axes as functions of Y, or otherwise, sketch Y as a function of t for a range of initial conditions.

[~15% marks]

(f) Comment on the behaviour of the population in the context of the biological meanings of the parameters α and β.

[~10% marks]

SECTION D

D7

Consider the non-linear system of equations given by:

$$\frac{dx}{dt} = xy^2 - 3xy$$

$$\frac{dy}{dt} = y - x^2 + x - 1$$

(a) Find and classify all equilibria.

[~30% marks]

(b) Sketch the phase plane, marking on clearly the nullclines, equilibrium points, direction field and sufficient trajectories to illustrate the behaviour of the system.

[~20% marks]

(c) Sketch the trajectory associated with initial conditions $(x_0, y_0) = (-0.5, 2)$ on the phase plane. In a new figure, sketch the corresponding graphs for *x*(t) and *y*(t) on the same set of axes.

[~25% marks]

(d) For the initial condition$(x_0, y_0) = (0, 2)$, sketch the graphs for *x*(t) and *y*(t) on the same set of axes. Confirm your answer by solving the differential equations analytically.

[~25% marks]

D8

In this question, we model the spread of rabies in a large population of domestic dogs. In the absence of rabies, the natural death rate of dogs is represented by parameter $\delta > 0$ and the population is managed by owners with a constant import rate $\alpha > 0$. Rabies is invariably lethal (there is no recovery), killing infected dogs with rate $\mu > 0$. We define $S(t)$ and $I(t)$ as the respective densities of susceptible and infected dogs, and model the transmission of rabies between dogs with rate $\beta > 0$. This leads to the following pair of differential equations:

$$\frac{dS}{dt} = \alpha - \beta\, S\, I - \delta\, S$$

$$\frac{dI}{dt} = \beta\, S\, I - \mu\, I$$

(a) In the absence of rabies, what is the stationary population density? You do not have to prove the stability of this equilibrium.

[~10% marks]

(b) Explain why the basic reproductive number of rabies in this model can be written as:

$$\mathcal{R}_0 = \frac{\beta\alpha}{\delta\mu}$$

[~10% marks]

(c) Calculate the null-clines and stationary points of the model, considering the cases where $\mathcal{R}_0 > 1$ and $\mathcal{R}_0 \leq 1$.

[~30% marks]

In the next questions, we assume that $\mu > \delta$ and $\mathcal{R}_0 = 2$.

(d) Show that the stationary point with endemic persistence of rabies is a stable focus.

[~25% marks]

(e) Sketch the phase plot (with S on the x axis and I on the y axis), including the direction field (with justification), and draw a trajectory starting from the introduction of small number of rabies cases in a susceptible population.

[~25% marks]

SECTION E

E9

The active site of enzyme EH contains a single acidic group that can dissociate to release a hydrogen ion, H^+. For catalysis to occur, the acidic group must adopt its negatively-charged dissociated form, in a species called E^-. The enzyme's substrate, S^+, bears a positive net charge and participates in the following reaction scheme:

$$EH \underset{k_2}{\overset{k_1}{\rightleftharpoons}} H^+ + E^-$$

$$E^- + S^+ \underset{k_4}{\overset{k_3}{\rightleftharpoons}} ES \xrightarrow{k_5} EH + P$$

(a) Acid-base reactions are typically very rapid, so the protonated and deprotonated forms of the enzyme will be in dynamic equilibrium with each other. Use the law of mass action to derive an equation that equates the rate of the deprotonation reaction with that of its protonation counterpart. Then, show how this generates two expressions for K_a, the equilibrium constant for the dissociation of EH, in terms of either rate constants or the concentrations of three distinct chemical species.

[~10% marks]

(b) Construct a molar balance equation for the total concentration of all enzyme species that are present at any time, $[E]_{TOT}$.

[~10% marks]

(c) Write down two differential equations that describe how the rates of change of [ES] and [P] vary as a function of time, t.

[~10% marks]

(d) If K_M is defined as $(k_4 + k_5)/k_3$, apply the steady state approximation to the enzyme-substrate complex and show that $[ES] = [E^-][S^+]/K_M$.

[~10% marks]

(e) Use your answers to parts ***(a)***, ***(b)*** and ***(d)*** to construct an expression for the current concentration of free dissociated enzyme, $[E^-]$, in terms of constants and the current concentrations of two ionic species.

[~10% marks]

(f) Use your answer to part ***(e)*** to show that the initial rate of product formation, v_0, obeys a rate law equation of the form:

$$v_0 = \frac{V_{MAX}^{app}[S^+]_0}{K_M^{app} + [S^+]_0}$$

where $[S^+]_0$ is the initial concentration of substrate. State any approximations you have made and explain how V_{MAX}^{app} and K_M^{app} depend on the values of other constants and the concentrations of specific species.

[~20% marks]

(g) If the H^+ ion can be described as an inhibitor, supply the appropriate name for this type of inhibition.

[~10% marks]

(h) Use your answer to part ***(f)*** to obtain simplified expressions for V_{MAX}' and K_M' in the two extremes when $[H^+]$ is very low and when $[H^+]$ is very high.

[~20% marks]

E10

A parent bird tending a brood of chicks must choose a (non-negative) level of effort *c* to invest in their care. A parent that invests effort *c* obtains an expected fitness payoff of

$$W(c) = \frac{c}{b+c} - kc$$

where the parameters *b* and *k* are both positive (with $bk \leq 1$). The first term in the above expression represents the parent's expected fitness gain from the current brood, and the second term the cost of care.

(a) Show that, for a single parent, the optimal level of caring effort, c_{single}, that maximises payoff is given by

$$c_{\text{single}} = \sqrt{\frac{b}{k}} - b$$

[~20% marks]

(b) What payoff does a single parent that invests this optimal level of caring effort obtain?

[~20% marks]

(c) Now suppose that two parents are tending the brood together and must each (independently) choose a (non-negative) level of caring effort. The payoff to a parent that invests effort c_{self} while its mate cares at level c_{mate} is given by

$$W(c_{\text{self}}, c_{\text{mate}}) = \frac{c_{\text{self}} + c_{\text{mate}}}{b + c_{\text{self}} + c_{\text{mate}}} - kc_{\text{self}}$$

Derive the optimal (or 'best-response') level of care for a parent, denoted $c_{\text{opt}}(c_{\text{mate}})$, that maximises its payoff given that its mate cares at level c_{mate}.

[~30% marks]

(d) Using your answer to *(c)*, derive a stable level of care c^* for which $c_{\text{opt}}(c^*) = c^*$, i.e. a level of care that is optimal for a parent when its mate invests that same level of care. How does c^* compare to c_{single}?

[~30% marks]