



COGMASTER

SECOND-YEAR INTERNSHIP MEMOIR

Self-sacrifice as a social signal

Julien Lie

supervised by

Jean-Louis DESSALLES

June 8, 2019

Contents

Declaration of originality	1
Declaration of contribution	1
I Pre-registration document	3
II Self-sacrifice as a social signal	12
1 Why do humans self-sacrifice for their groups?	13
1.1 Self-sacrifice	13
1.1.1 Prevalence and characteristics of self-sacrifice in humans	13
1.1.2 Identity fusion	14
1.2 Existing ultimate (biological) explanations for self-sacrifice . .	15
1.2.1 Self-sacrifice does not need (more) explaining	16
1.2.2 Kin selection	16
1.2.3 Group benefit; cultural selection	19
1.3 Hypothesis: self-sacrifice as a social signal	20
1.3.1 Biological signaling	20
1.3.2 Social signaling and human prosocial behavior	22
1.3.3 Social signaling and high-cost prosocial behavior	23
1.3.4 Self-sacrifice: a two-tier social signaling model	24

1.3.5	General predictions	26
2	Methodes used during the internship	28
2.1	Objective and methods	28
2.1.1	Objective	28
2.1.2	Methods	29
2.2	Computer simulations	29
2.2.1	Evolife	29
2.2.2	“Exogenous” base script	30
2.2.3	Actual script structure	31
2.2.4	Patriots and non-patriots	33
2.2.5	Main variable parameters	35
2.2.6	Other fixed parameters (at least in a first stage)	35
2.3	Predictions	36
3	Results	38
3.1	Output	38
3.1.1	Typical simluations	38
3.1.2	Remembered Heroes	38
3.2	Exogenous model: simulation outputs	39
3.2.1	Main results	39
3.2.2	Influence of <i>ReproGainsThreshold RGT</i>	40
3.2.3	Variation with <i>SacrificeHeredity h</i>	42
3.2.4	Variation with <i>ReproductionRate r</i>	42
3.3	Exogenous model: mathematical proof of concept	43
3.3.1	Simplifying assumptions and characterization of equilibriu	43
3.3.2	Necessary condition for an ESS	44
3.3.3	Mathematical characterization of the ESS	45
3.3.4	Brief Discussion	45

3.4	Two-tier model simulation outputs	46
3.5	Two-tier model: mathematical analysis	46
3.5.1	Self-sacrifice	46
3.5.2	Honoring and social score	47
3.5.3	Equilibrium when $t \leq \frac{F}{FC}$	47
3.5.4	"Honest" equilibrium when $t \geq \frac{2F+P}{DC}$	48
3.5.5	Dishonest equilibrium when $t \geq \frac{2F+P}{DC}$	49
3.5.6	$\frac{F}{DC} < t < \frac{2F+P}{DC}$	50
3.6	Second sacrifice model	50
3.7	Limitations	50
4	Perspectives	51
A	Python scripts	52
A.1	Exogenous.py	52
A.2	Sacrifice.py	64
B	Mathematical demonstrations	72
B.1	Exogenous model	72
B.1.1	RemTH	72
B.1.2	$\beta(p)$	72
B.1.3	$R_+(A)$	73
B.1.4	ESS	75

Declaration of originality

- Déclaration d'originalité

Les stages de M2 Recherche doivent être de véritables travaux de recherche originale, pouvant aller au-delà des connaissances déjà publiées (sans préjuger des résultats obtenus). Afin de clarifier cet aspect, chaque étudiant doit faire figurer, en première page de son mémoire, une déclaration d'originalité spécifiant en quoi le travail présenté va au-delà de ce qui est déjà connu.

pour les travaux expérimentaux préciser en quoi les expériences effectuées diffèrent de celles déjà publiées, et en quoi elles pourraient potentiellement permettre de contribuer à résoudre une question non encore élucidée (sans préjuger des résultats obtenus) pour les travaux théoriques préciser en quoi les idées, modèles ou théories proposés vont au-delà de ceux qui sont publiés, et ont éventuellement un pouvoir explicatif supérieur

Declaration of contribution

Directly participated in this work: Jean-Louis Dessalles, myself.

- Definition of the scientific question: Jean-Louis Dessalles, myself;
- Bibliographical research: Jean-Louis Dessalles, myself. I also greatly benefited from related work evoked during Cogmaster classes, notably CA9 and CA11, given by Jean-Baptiste André, Nicolas Baumard and Coralie Chevalier;
- Choice of the general approach: Jean-Louis Dessalles;
- Choice of the specific methodology: Jean-Louis Dessalles;
- Development of the methodology (Evolife): Jean-Louis Dessalles;
- Development of the specific scripts: myself. I am also greatly indebted to the Stackoverflow community (technical forum), as well as to Christophe Pradier's PCBS class (introduction to Github notably);
- Data analysis: myself;
- Mathematical analysis: Jean-Louis Dessalles, myself;
- Interpretation of results: Jean-Louis Dessalles, myself;
- Memoir writing, production of figures: myself (and Stackoverflow);
- Comments on the memoir: Jean-Louis Dessalles.

Part I

Pre-registration document

Julien Panis-Lie
Self-sacrifice as social signal

supervisor: Jean-Louis Dessalles
Département Informatique et Réseaux
Télécom Paris

Defense target date: June
Language: English

Reviewer suggestion (non CP): Jean-Baptiste André
Reviewer suggestion (CP): Nicolas Baumard or Coralie Chevallier

Self-sacrifice as social signal

Modeling the motivations underlying an extreme example of prosocial behavior

1. Introduction.....	2
2 Methods.....	3
2. 1. Implementation of the model.....	3
2. 2. Script structure.....	3
2. 3. Important parameters.....	4
2. 4. Interpretation.....	5
2. 5. Possible refinements of the model.....	5
References.....	6

1. Introduction

Throughout history, humans have been willing to lay down their lives for the sake of their groups (Whitehouse, 2018). Whether admired as heroes or reviled as terrorists, individuals who engage in such extreme pro-social behavior tend to be relatively well-off, educated and to display no appreciable psychopathology (Atran, 2003) – underscoring that self-sacrifice cannot simply be viewed as the result of unadaptive miscalculation or proximal causes (e. g. the contents of a particular ideology).

Could self-sacrifice therefore have a biological function? This project purposes to investigate the potential biological motivations that may underlie this behavior. Even though collective benefit is the displayed (moral) motivation for self-sacrifice, biological motivation will be understood in terms of benefits to the individual (Williams, 1996).

To do this, we will attempt to model these biological motivations, using a *social signaling framework* (Dessalles, 2014). Social signals are a specific case of costly signals (Zahavi, 1975; Grafen, 1990) whose purpose are to attract friends. In contexts where the signaled quality correlates with the fitness of friends, such a quality may be in social demand; when, in addition, the potential benefits in terms of increased social status upset the costs, social signaling can be expected to emerge.

This explanation cannot hold for self-sacrifice, however, since signalers would not survive to enjoy the advantages of their new status. We propose to add an additional hypothesis to our theoretical framework: *that social status be in part heritable for our species*, as the high status of one individual can raise that of every member of his or her family (Service, 1971). In a context where a martyr's family members are in high social demand, this could theoretically suffice to make self-sacrifice evolutionary stably – as long as it remains a low-frequency behavior (the fewer the heroes, the higher their status).

We are thus left with the need to find a plausible context where a martyr's family may be in high social demand. This is not immediately obvious, as the quality signaled by individuals who engage in self-sacrificial behavior need not be heritable. We propose to consider a context akin to *inter-group conflict*, where it may be crucial to ensure that one's friends carry no sympathy for the enemy (e. g. to avoid betrayal). In such a situation, *commitment to the group* is a highly desirable quality to have in friends. One way for (alive) individuals to signal this quality may be to honor the fallen martyrs, for instance by engaging in conspicuous ceremonies in their name. If and when such costly second-order signaling is evolutionary stable, honoring could entail indirect benefits for the family of martyrs, meaning that self-sacrifice may itself emerge – the two signals being expected to be mutually reinforcing, as would-be-patriots need martyrs to signal their commitment to the group.

Our main hypothesis, which we venture to investigate, is therefore the following: in a biological population where social status is in part heritable, and which is engaged in a context akin to inter-group conflict, self-sacrifice and honoring may emerge as first-order and second-order signals of individuals' commitment to the group, or patriotism. Using computer simulations, our project will study the conditions (if any) where these two signals may emerge (e.g. cost and probability of betrayal).

2. Methods

2. 1. Implementation of the model

The script is written in Python, and set in the Evolife framework, developed by Jean-Louis Desselles to study various evolutionary phenomena (<https://evolife.telecom-paristech.fr/>). Evolife has been used to study social signals in general (Desselles, 2014) and to model language as a particular social signal (Desselles, 2017).

Evolife is based on a genetic algorithm. An individual's behavior is controlled by a binary vector (genome). Individuals live, reproduce sexually, and gain points in a (yearly) life game. Two modes of selection can be implemented:

- *ranking*: individuals are ranked following the points they have obtained, and are granted a number of potential children that is an increasing (non-linear) function of that rank ;
- *differential death*: individuals are granted life points (related to their yearly points) which protect them from life hazards, thus increasing their life expectancy (and reproductive opportunities).

As detailed below, the proposed script exploits this dichotomy in its treatment of the envisioned first-order and second-order signals. Evolife is modular in structure; the script would constitute a scenario implemented inside of Evolife. Source files (<https://github.com/jlie10/SelfSacrifice>) can be found on GitHub – on the “Internship” branch (“Master” being kept for a pending class validation).

2. 2. “Exogenous” base script

Before implementing the actual script, a simplified “exogenous” version was implemented (Exogenous.py). In this version, the social value of self-sacrifice is fixed exogenously, and depends on an ‘Admiration’ parameter. In order to make the link to the second script more obvious, we supposed that total admiration for heroes was equal to ‘Admiration’ times the number of individuals that don’t sacrifice – seeing the low final yearly frequency of self-sacrificial behavior (see under), a truly fixed version (where total admiration does not depend on the number of heroes) would yield similar results.

This script serves as a basis for the full script. Self-sacrifice is treated as a genetically controlled trait: individuals in the scenario are endowed with a ‘SelfSacrifice’ gene (8 bits) whose relative value corresponds to an individual’s probability of engaging in self-sacrificial behavior in a given year.

A heroes’ sacrifice benefits his/her descendants. Each year, admiration for heroes “spills over” to the rest of the population. An individual’s ‘share’ in this allocation depends on their descendant’s heroism: each time a parent engages in self-sacrificial behavior, his/her children gain 1 share point. Shares are inherited, as individuals are born with the average of their parent’s share times a ‘SacrificeHeredity’ parameter, situated between 0 and 1. This parameter is generally kept under 1 (typically 0.5), to avoid a situation where individuals end up strongly related (as a given heroes’ family invades the population).

The emergence and stabilization of self-sacrifice is studied following a ranking mode of selection. Admiration points indirectly gained by individuals are converted into reproductive points, by taking the integer part of these points divided by a ‘ReproGainsThreshold’ parameter (typically fixed at 10).

Keeping other parameters at their typical values (see after), self-sacrifice is, unsurprisingly, shown to

depend on ‘Admiration’ (relative to ‘ReproGainsThreshold’): when this parameter is too low (e.g. under 5 – tentatively), this gene’s value remains at 0 across the population; when the parameter is augmented, the equilibrium value of the gene grows with it – although it is quickly capped at 4-5 % (which, for a small population of 200, entails an average of 10 sacrifices per ‘year’ - which is far from negligible).

Full statistical analysis of this behavior has yet to be performed. The idea, in the following scenario, is to see if, in conditions where we would expect (alive) individuals to pay tribute to heroes on a scale comparable to the previous values of ‘Admiration’, self-sacrifice emerges as a stable strategy (as measured by gene value across the population, and as compared to results obtained in similar conditions for this base scenario).

2. 2. Actual script structure

This script (Sacrifice.py) builds on the previous one. The idea is to see if admiration for heroes, and therefore self-sacrificial behavior, can emerge “endogenously”, in a context akin to inter-group conflict, as described previously. The previous dynamic and ranking mode of selection is conserved for self-sacrifice, while other genes (see under) follow a differential death mode of selection (as points gained or lost in the social interactions detailed under are later translated into life points).

In addition to the ‘SelfSacrifice’ gene, individuals are born with a ‘Patriotism’ phenotype – which, for the sake of simplicity is equal to 1 or 0. Individuals are also endowed with a ‘Demand’ gene, as well as a ‘Patriot’ and a ‘NonPatriot’ gene, which respectively control their yearly investment in honoring heroes when they are patriots (1) or not (0).

When there are heroes to honor, this (costly) investment becomes a visible signal of their commitment to the group (patriotism). Individuals select their friends according to the potential signalers they encounter (if there are no heroes, every individual signals at 0) and their ‘Demand’: any individual who signals above this value may be accepted as a friend (one interaction yields one new friend at the maximum).

Friendship is assumed to be mutually beneficial: an individual with a low ‘Demand’ will thus increase his/her chances of gaining from friendship. However, some ‘NonPatriot’ individuals may be traitors: befriending them ends up being extremely costly, as they betray their friends in the final stage of the year, to their own gain. Thus, in uncertain conditions where being betrayed is a probable and costly outcome, individuals with high ‘Demand’ may fare better.

In conditions where honoring is an honest signal of one’s future (absence of) betrayal, ‘Demand’ and ‘Patriot’ may thus potentially co-emerge, which may then (if ‘Patriot’ is high enough) allow ‘Self-sacrifice’ to emerge. This could be the case if non-patriots honoring potential is capped: because non-patriots are already committed to another group, or themselves, they may not have the time or resources to invest as much as patriots are able to.

2. 3. Important parameters

2. 3. 1) Main variable parameters

- ‘NbTraitors’: probability that a non-patriot actually ends up being a traitor. This is the main

parameter, which will vary between 0 and 100 ;

- ‘MaxOffer’: maximum investment that non-patriots can afford in honoring (out of 100). Fixed at 50 in an initial stage, although exploring other values – including 100 – will be useful (see after) ;
- ‘FriendshipValue’: value gained from being friends with a non-traitor – fixed at 10 in an early stage ;
- ‘JoiningBonus’: value gained from having been accepted as friend by someone else – fixed at 10 in an early stage ;
- ‘Judas’: what a traitor gains from betraying you – fixed at 20 in an early stage ;
- ‘DenunciationCost’: cost of being betrayed – fixed at 100 in an early stage.

Individuals start off the year with 100 points. The previous values for social points (last four parameters) were obtained via reasonable estimates from this scale, and after numeric simulations of honoring in isolation (see Honor.py). Given all these parameters, ‘NbTraitors’ controls the probability of betrayal. In order to differentiate with the cost of betrayal, the latter two parameters will also be made to vary around the previously specified values.

In this simplified version with only two levels of patriotism, ‘MaxOffer’ creates an arbitrary cutoff, above which a signal can be considered as honest. For this reason, ‘MaxOffer’ will be largely kept constant at the arbitrary 50, although looking at 100 could be useful to understand what happens when honoring is not honest.

2. 3. 2) Other fixed parameters (at least in a first stage)

With respect to the (expected) first-order signal:

- ‘ReproGainsThreshold’: corresponds to the value in terms of points acquired through honoring of one’s descendants that yields an additional reproductive unit (for ranking) – kept at 10 ;
- ‘Selectivity’: controls how the degree of selection by ranking (how much expected number of offspring increases with rank) ;
- ‘SacrificeHeredity’: controls how shares are inherited from parents (see 2.1) – fixed at 50 in a first stage ;
- ‘ReproductionRate’: expected rate of children left each year. Fixed at 30%, a relatively high value, as (potential) equilibrium value for probability of self-sacrifice can be shown to be inferior to reproduction rate.

With respect to the (expected) second-order signal:

- ‘HonoringCost’: percent of investment in honoring that translates into cost – fixed at 100 ;
- ‘SelectionPressure’: maximum number of life-points that can be earned in a round – fixed at 6 (meaning that individuals with the highest scores have to be randomly selected 7 times before enough other individuals are selected between 1 and 6 times, in order to die that turn) ;
- ‘EraseNetwork’: indicates whether individuals’ friendship networks are reinitialized each year. Fixed at 0 (False), mainly in order to avoid having to largely increase ‘Rounds’ and ‘NbInteractions’ (and hence computation time).
- ‘Rounds’: number of times friendship-forming interactions are launched in the population – fixed at 10 ;
- ‘NbInteractions’: number of such interactions per round, where one randomly chosen individual may accept a maximum of one friend – fixed at 100 ;
- ‘SampleSize’: size of the random sample of (live) people an individual thus interacts with – fixed at 5 ;
- ‘MaxFriends’: maximum number of friends one can hold – fixed at 10. Friendship bonds are assumed

to be symmetrical here ;

Genetic and general population parameters:

- 'AgeMax': age after which individuals automatically die – fixed at 40 ;
- 'PopulationSize': fixed at 1000 ;
- 'MutationRate': fixed at 5 per 1000 ;
- 'NbCrossover': number of crossovers occurring during sexual reproduction – fixed at 1 ;
- 'GeneLength': fixed at 8, weighted bits (meaning that each gene value corresponds to a value between 0 and 2 to the power of 8 minus 1)
- 'AgeAdult': age before which reproduction is impossible – fixed at 0

It is anticipated that keeping 'EraseNetwork' at 0 and 'Rounds' and 'NbInteractions' relatively low is equivalent to re-initializing networks with large number of interactions – although this should be checked.

In addition, a 'DiffentialCosts' mode for honoring could be implemented, where dishonest signalers (non-patriots) would face a premium, rather than a capped offer.

2. 4. Predictions

Our main prediction is that, under plausible conditions corresponding to the default parameter values detailed above, honoring and self-sacrifice will co-emerge as second and first-order signals of patriotism (inability to betray). We expect this to be largely controlled by 'NbTraitors' (probability of betrayal).

When and if self-sacrifice emerges, we do not expect its probability to be high, for logical reasons (as gains in status decrease when there are too many perpetrators) as well as reasons relating to previous simulations (we do not expect equilibrium values to exceed that of the corresponding "exogenous" situations). In addition, we do not expect honoring to be stable in the long-term in the absence of heroes (as we envisioned it as a second-order signal based on the first-order signal).

More specifically, keeping all other parameter values constant, we expect:

- that for 'NbTraitors' equal to 0 or sufficiently small, 'Demand' should stay at 0 or non-significantly higher (as friendship is risk-free), thus inducing 'Patriot' and 'NonPatriot' to stay low, preventing the emergence of self-sacrificial behavior ;
- that when 'NbTraitors' is high enough for there to be a selective pressure for having 'Demand' above 'MaxOffer', honoring is a potentially honest signal. If, in addition, the benefits to patriots outweigh the costs, then honoring may emerge at such a level. Seeing as we chose 'MaxOffer' at a level which, in the "exogenous" script (for 'Admiration') was sufficient for self-sacrifice to emerge and stabilize, we that, provided that honoring can emerge, that self-sacrifice emerges as well ;
- if we vary 'DenunciationCost' instead of 'NbTraitors', we expect similar dynamics ;
- if we decrease 'JoiningBonus' under a certain threshold, attracting friends through signaling patriotism should become overly costly (risky), preventing any non-null behavior at equilibrium.

We expect 'MaxOffer' to play a decisive role, as it should correspond to the potential equilibrium state – when it is attainable. Testing 'MaxOffer' at 100 would be a way of seeing the consequences of dishonest signaling: under normal conditions we would expect that 'Demand' then rise to the maximum (as it is impossible to select patriot friends), precluding any signaling from emerging.

However, if ‘Judas’ is high enough, signaling may paradoxically remain interesting for non-patriots and we would expect a dishonest signaling equilibrium to emerge.

2. 5. Limits and possible refinements of the model

‘MaxOffer’ thus imposes a paradoxical arbitrary limit: if it is too low, we would not expect self-sacrifice to emerge, as honoring of heroes by patriots should not yield enough potential benefits for heroes’ children (see 2.1).

This is further motivation to explore a “differential costs” option, whereby dishonest signaling is assumed to be costlier than honest signaling (e.g. for the same type of reason, as non-patriots opportunity costs are higher, as they are engaged elsewhere / have other opportunities to invest in themselves...). This could allow for less straightforward results, as we would expect both signals to emerge above a certain threshold (for the ‘DishonestPremium’) and not under, without this threshold being immediately obvious.

However, we expect both treatments to be mathematically equivalent, as the poverty of our (expected) ability to modify final outcomes by varying parameter values derives from the poverty of the model itself. In a second-stage, the model could be refined:

- by allowing for more categories of patriotism, and the corresponding genes ;
- or, better yet, by allowing for continuous (or a large number of discrete values for) patriotism, and replacing genes by social learning of investment in honoring.

2. 6. Interpretation

The immediate outputs of the simulation are average (and individual) levels of gene values over time. If, as predicted, average value for the gene controlling self-sacrificial behavior stabilizes at a non-null value, and if the various parameters (and ‘Honoring’ genes) influence this according to our predictions, then we will have shown that biological motivations underlying self-sacrifice are a theoretical possibility. We would then study stability of the model itself over a certain range of parameters.

As detailed above, interpretation is however compounded by the current simplistic form of the model.

References

- Atran, S. (2003). Genesis of Suicide Terrorism. *Science*, 299(5612), 1534-1539.
<https://doi.org/10.1126/science.1078854>
- Dessalles, J.-L. (2014). OPTIMAL INVESTMENT IN SOCIAL SIGNALS: OPTIMAL INVESTMENT IN SOCIAL SIGNALS. *Evolution*, 68(6), 1640-1650.
<https://doi.org/10.1111/evo.12378>
- Dessalles, J.-L. (2017). Language: The missing selection pressure. *arXiv:1712.05005 [physics, q-bio]*. Consulté à l'adresse <http://arxiv.org/abs/1712.05005>
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144(4), 517-546.
[https://doi.org/10.1016/S0022-5193\(05\)80088-8](https://doi.org/10.1016/S0022-5193(05)80088-8)
- Service, E. R. (1971). *Primitive social organization: An evolutionary perspective*. Random House, Toronto, CAN.
- Whitehouse, H. (2018). Dying for the group: Towards a general theory of extreme self-sacrifice. *Behavioral and Brain Sciences*, 1-64. <https://doi.org/10.1017/S0140525X18000249>
- Williams, G. C. (1996). *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton, NJ: Princeton Univ. Press.
- Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, 53(1), 205-214. [https://doi.org/10.1016/0022-5193\(75\)90111-3](https://doi.org/10.1016/0022-5193(75)90111-3)

Part II

Self-sacrifice as a social signal

1 | Why do humans self-sacrifice for their groups?

1.1 Self-sacrifice

1.1.1 Prevalence and characteristics of self-sacrifice in humans

Throughout history, humans have been willing to lay down their lives for the sake of their groups “altruistic suicides”, encompassing any behavior which will *necessarily result in death*, in which individuals engage knowingly, *in the name of a group and/or its ideology*. Throughout this paper this will be referred to as extreme *prosocial* self-sacrifice, as it involves extreme costs to the self (death) intended for the benefit of others – or simply self-sacrifice. Perpetuators of such acts will alternatively be referred to as "(would-be) martyrs" or "heroes", without any distinction between the two.

Early Christian martyrs (Durkheim, *ibid*), the 300 Spartans at the Battle of Thermopylae terrorists in recent decades

These examples underscore that such self-sacrificial behavior may be related to (perceived) threat to the group, perhaps notably in a context of intergroup conflict. Intergroup violence is a widespread and persistent feature of sapiens’ environment during the Pleistocene, with potentially far-reaching impact on our mortality making it likely that mechanisms related to such a situation would have been selected in our species.

In contrast however, the sectarians of Amida in the early nineteenth century (Durkheim, *ibid*) appear to give up their life outside of any such context¹ in *public* displays of piety, their *memory* being held in great reverence

¹ Although it is hard to rule out (and not specified here) that they may be motivated

by members of the crowd. Martyrs engaged in intergroup conflict also gain posthumous celebrity within the group, which appears to *benefit their families*

Self-sacrifice	
Definition	In the name of a group and/or ideology Results in death Voluntary behavior
Context	Recurrent historical fact For a variety of groups and/or ideologies Seems tied to intergroup conflict
Outcome	"Martyrs" are often revered and memorialized This may benefit their family

Table 1: Characteristics of self-sacrifice (tentative summary).

1.1.2 Identity fusion

For *identity fusion*, a visceral sense of oneness with the group Two pathways can lead to enduring identity fusion: perceptions of shared biology, or *psychological kinship*, as well as intense collective experiences, including the horrors of frontline combat or participation in potentially extremely *painful rituals*. Highly “fused” individuals appear to take threats to the group personally, and, in extreme cases, may be motivated to lay down their life for said group. While the difficulty of evaluating actual would-be martyrs’ motivations is compounded by practical and ethical issues, one can note that identity fusion is correlated with expressions of support for martyrs and even stated willingness to give up one’s own life to defend the group

Identity fusion accounts for many of the characteristics of self-sacrifice as summarized in **Table 1**, including its relation to conflict and its stated objective (defense of the group). In addition, it allows to place willingness to lay down one’s life for a group on a spectrum and connect it with two other associated social mechanisms – extreme rituals (e.g. intense initiations, (or other family-like ties) in public discourse.

by a perceived threat to their sect.

Alternative explanations put a more direct emphasis on certain elements evoked above: ideology familial ties also play a causal role.

From an evolutionary standpoint however, these explanations only beg the question. Beliefs, psychological motivations and social mechanisms, whether or not they are integrated under the concept of identity fusion, are immediate proximate causes for behavior, which cannot account for the evolution of self-sacrifice in our species. Seeing its recurrence in historical records, it seems defensible to assume, as we will throughout this memoir, that self-sacrifice is an evolved capacity (or at least that some of its major underlying proximal causes are), and can be studied from a Darwinian perspective (see **Section 1.2** for varying counterarguments; and justification for (**H1**)).

Hypothesis 1 (H1): *Self-sacrifice can be understood as an evolved capacity.*

At first glance, self-sacrificial behavior and its underlying motivations constitute a true biological puzzle, since dying is obviously not a good way of passing one's genes. Following (**H1**), this paper aims to better understand why self-sacrifice, as characterized above, exists, by taking an evolutionary outlook (**Q1**). In addition, this paper hopes to provide a potential explanation that takes into account all of self-sacrifice's arguably important features, as summarized in **Table 1 (Q2-4)**.

Question 1 (Q1): *What are the ultimate causes of self-sacrifice for the sake of the group, which may account for the evolution and maintenance of such behavior in humans?*

Question 2 (Q2): *Why should self-sacrifice be related to intergroup conflict?*

Question 3 (Q3): *Why are martyrs often revered and memorialized?*

Question 4 (Q4): *Why should a martyrs' self-sacrifice benefit their family?*

1.2 Existing ultimate (biological) explanations for self-sacrifice

Explanations for self-sacrifice usually invoke maladaptive behavior (pathology, miscalculation...), kin selection or group benefit – or, a mixture of the three.

1.2.1 Self-sacrifice does not need (more) explaining

For some authors, self-sacrifice (particularly in the modern form of suicide terrorism) is caused by extreme religious views and/or pathology. This does not square however with the extent of such behavior, which is also displayed by individuals one would more intuitively link to patriotic or nationalist beliefs (Spartans at Thermopylae, kamikaze during World War II) - which prompted us to take hypothesis (**H1**) above.

More importantly, the persistence of self-sacrifice throughout our history suggests that pathology may not be a sufficient explanation. If self-sacrifice is maladaptive, seeing its huge costs to individual fitness, *why did such behavior evolve in our species (Q1)?* In addition, contemporary studies of individuals who engage in such behavior suggest they have no appreciable psychopathology and are as educated and well-off as surrounding populations.

A related argument is that such behavior should not be understood as functionally self-sacrificial. The biological *function* of a behavior is an effect of that trait that causally explains its evolution and persistence in a population.

As outlined in **Section 1.3.3**, heroism may have a social function which allows to account for the evolution of heroic behavior particularly in times of intergroup conflict. Such heroic behavior may sometimes result in death, manifestly in favor of the warring group. In such an interpretation, self-sacrifice is thus accidental, and/or borne from miscalculation.

The distinction between risking one's life and laying it down is not always clear-cut. According to one estimate of surviving an act worthy of a British Commonwealth Victoria Cross medal, making it hard to decide whether such acts should be characterized as heroic or self-sacrificial. For cases such as the examples given above, which involve long-term planning (e.g. Muslim suicide terrorists) and repetitive actions that can only realistically result in death (e.g. Christian martyrs), it seems hard to avoid the latter qualification – unless one factors in severe and dire miscalculation. This explanation is thus similar to the previous “pathological” one; the same objections can be made.

1.2.2 Kin selection

Throughout the natural world, one example of prosocial behavior abounds: parental care. Costly behavior in favor of children (and more generally, kin) is adaptive if, from the standpoint of genes, costs are outweighed by benefits, as captured by Hamilton's rule smaller than benefits b to a kin times coefficient

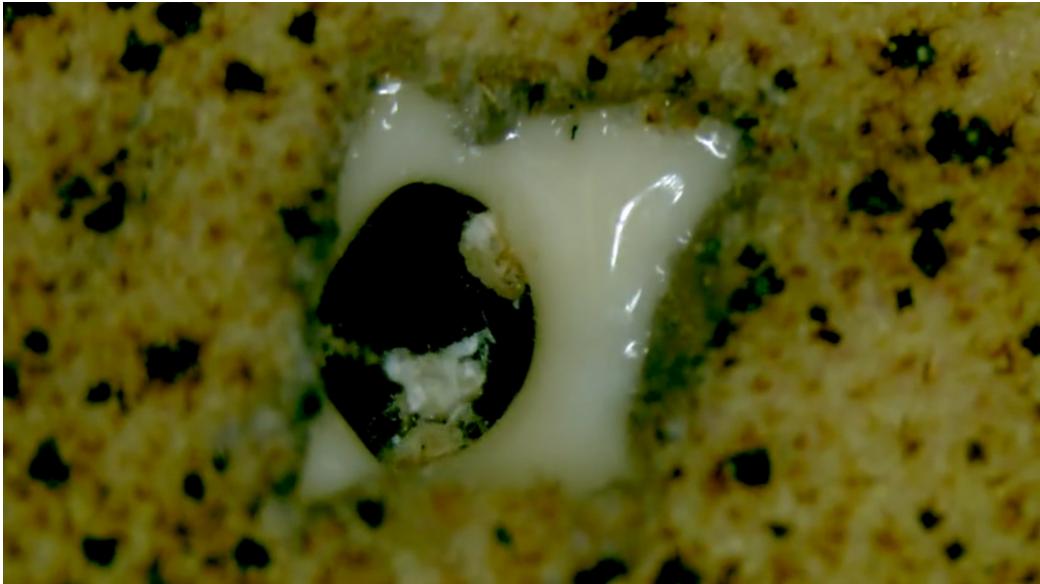


Figure 1: Soldier aphids using their body fluid to repair their gall. Kutsukake et al., via lemonde.fr.

of genetic relationship r , then, on average, a gene that favors said behavior will leave more copies than a gene that does not. At the individual level, this is captured by the concept of *inclusive fitness*.

$$\text{Hamilton's rule: } b * r > c$$

Hamilton's rule leaves the door open to even self-sacrificial behavior in favor of kin. Thus, in some (rare) examples, such as certain spiders mothers are systematically eaten by their offspring (and let them do it) at the end of their incubation period, a behavior which is sufficiently beneficial to the long-term prospects of the (tens of) individuals in the clutch to have evolved. Beneficiaries need not be restricted to immediate offspring: certain aphids and other social insects may “self-explode”, in order to plaster over gaps in the nest (gall) with their body fluid, to the benefit of the entire colony

Whitehouse and Lanman the evolution and maintenance of identity fusion – and self-sacrifice, in extreme cases of intergroup conflict. In such an interpretation, both pathways leading to fusion (perceived biological relatedness and shared experiences) and their corresponding social mechanisms (evoking “brotherhood”, rituals...) come down to the same fundamental issue: detecting your genetic kin, and motivating potentially extreme behavior in their favor, when the situation demands it (conflict). This will be referred to

as “*kin fusion*”.

Kin fusion rests on the debatable assumption that ancestral warring groups were composed of close genetic relatives. Of course, Whitehouse and Lanman do not anticipate that such genetic proximity should be comparable to that in an aphid gall², but merely significant enough for similar kin self-sacrifice to remain a theoretical possibility. Studies of current hunter-gatherer groups closely related.

In any case, even if one accepts that self-sacrifice through fusion may evolved to favor ancestral groups of kin, the question remains: *why would such behavior be maintained (Q1)?* In the historical examples given, groups who are intended to benefit from self-sacrifice (e.g. national or religious community) are too large to be genetically related. From an evolutionary standpoint, feeling fused to such a large group to the point of self-sacrifice thus constitutes an extremely costly mistake³ – which brings us back to the arguments developed in the previous section.

Another way of complementing – or replacing – this explanation revolves around group and/or cultural selection, as detailed in the next section. Alternatively, Blackwell (2008) notes that, for Palestinian suicide fighters, self-sacrifice comes with material gains to the family. In certain economic situations, self-sacrifice may thus increase an individual’s inclusive fitness. This is similar to the arguments developed in **Section 1.3**, although Blackwell’s model does not specify how such a situation may come to be – in other words, he does not address question (**Q4**).

²Contrary to us, social insects reproduce via eusocial division of labor (a colony descends from one or a small number of “kings” or “queens”) or cloning. Two random individuals taken in a colony may thus be very closely related (tentatively, r measurable in tens of percentage points) – considerably more so than two random hunter-gatherers taken in a typical unit of over 100 individuals. If however, on average, these two individuals are significantly more related than two individuals in two different groups, self-sacrifice may still theoretically emerge, provided benefits to kin are large enough.

³Evolutionary theory can account for such “mistakes”, through the concepts of *evolutionary mismatch* and *proper* and *actual domain*. However, contrary to classical examples pertaining to the latter distinction (e.g. cultural representations of faces, part of our facial-recognition module’s actual domain, but not its proper domain, for which it was selected), self-sacrifice for a group (much) larger than kin comes with significant evolutionary cost. And, in contrast to classical examples of mismatch (e.g. appetite for sugar in countries where (refined) sugar is plentiful), self-sacrifice for groups of non-kin seems to have been around for a much longer period (as long as we have records, at least).

1.2.3 Group benefit; cultural selection

Another family of explanations starts from self-sacrifice's stated objective: collective benefit. Groups comprising prosocial individuals should fare better than groups of egoists at the collective level – offering another perspective on insect eusociality

Many authors object to the idea that natural selection should (also) occur at the level of the group, as, in practice, human collective dynamics seem not to verify its axioms (They argue that the fundamental level for natural selection is genetic: what determines the evolution of a heritable behavior is, all else being equal, the number of copies that genes³ controlling it leave in the next generation is merely an approximation, which can be made in numerous cases because of how inter-related the fate of an individual's genes are – a condition which does not seem to be met at the collective level⁴. At the individual level, purely prosocial behavior (with no supplementary benefit to the individual with respect to others) is a losing strategy, as illustrated by the tragedy of the commons – and should therefore be counter-selected.

In particular, explanations for self-sacrifice in terms of collective benefit make a questionable assumption: that groups may face threat of complete annihilation. Orbell and Moriwaka larger coalitions comprising non-related individuals, while Whitehouse et al. content that fusion and self-sacrifice can evolve directly for groups of non-kin, provided prosocial behavior is conditioned on past shared experiences E_j (by group j). Their two-tier model (individual and group levels) includes the same assumption, as groups performing badly (poor relative success P_j) increase their chances of being completely replaced by the offspring of another group, as determined by probability S_j (see under)⁴. Yet, from the standpoint of genes, prehistoric group extinction seems highly unlikely: even if the entirety of its (predominately male) fighting force is massacred, civilian women may survive to join other groups and/or be subject to rape, as is recurrent in such conflicts

$$\text{Group survival probability (Whitehouse et. al): } S_j = hE_j + (1 - h) * P_j$$

Another related explanation revolves around the idea of cultural selection: that cultural objects may follow a process akin to Darwinian natural selection evolved to exploit the previously described propensity to fuse and potentially self-sacrifice for kin (Orbell & Moriwaka, 2011; Swann et al., 2012; Whitehouse, 2018). Such an explanation could be more robust to the previous

⁴ $0 \leq h \leq 1$ is a constant in their model which determines the relative importance of past shared experience.

criticism: it may be less debatable to suggest that norms or institutions can disappear, although this neglects the fact that such cultural elements cannot exist purely outside of individuals' minds Darwinian axioms – in particular, cultural transmission seems far from random and culture does not appear to exhibit inheritance in the strict sense (

Additionally, from a purely individual standpoint, self-sacrifice in response to such a potentially selected norm or institution remains an evolutionary mistake, as argued above.

1.3 Hypothesis: self-sacrifice as a social signal

1.3.1 Biological signaling

Questions (**Q1-4**) were studied during the internship following a signaling framework. According to the evolutionary theory of costly signaling natural selection may, under certain conditions, lead to waste at the individual level (a handicap). Zahavi's handicap principle has helped explain counterintuitive phenomena across the natural world, from the brightness of male plumage in certain bird species whereby certain preys (e.g. gazelles) will jump up into the air upon predator encounter, apparently making them easier to catch

A typical signaling model involves senders (signalers) and an audience (receivers) and can be grounded in evolutionary game theory. Senders vary in some specific unobservable *quality* of interest to the audience (note that without variation, there is no need for signals). They may advertise this quality to their audience, which may infer actual quality from these signals and base subsequent choices on these inferences (e.g. mate selection or prey pursuit – see under). Under reasonable mathematical assumptions⁵, a signaling equilibrium consisting in a pair of evolutionarily stable strategies – or ESS – for senders and receivers, can be shown to exist. In other words, if one assumes that the strategies followed by senders and receivers – signaling at a certain level and inferring quality from signals – are biologically encoded and heritable, then natural selection can lead to a non-trivial signaling equilibrium which is resistant to invasion by mutants (alternative strategies).

Conversely, given such an ESS pair, one can deduce: that signaling should be *honest*, meaning that advertised levels should reflect actual quality; that

⁵Honesty, cost, and increasing cost for males of lesser quality – as well as more technical elements

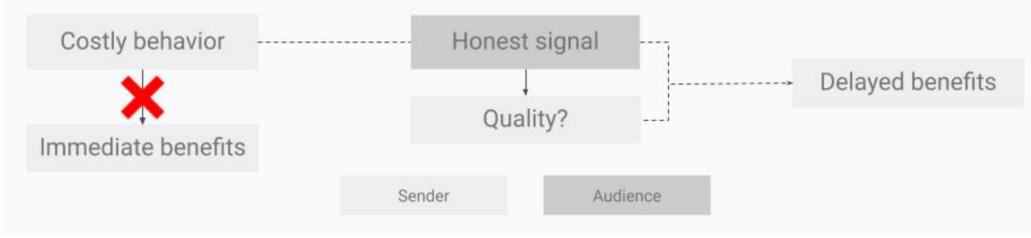


Figure 2: Signaling theory helps explain the emergence of costly behavior that is not followed by immediate benefits. Competition between senders helps guarantee signal honesty, and thus their relevance to the audience.

signalers should bear a fitness cost (handicap); and that said cost should be higher for senders of worst quality ethology: if one observes a situation where individual animals may be understood to be signaling a quality to an audience, then these signals should be honest and costly⁶. Honesty is key: if a signal is dishonest, then using it to infer quality would be sub-optimal – meaning that the corresponding strategy pair cannot be an ESS. This framework does not therefore apply to cases of dishonest animal communication.



(a) Gazelle (sender)



(b) Cheetah (audience)

Figure 3: Gazelle stotting. Images extracted from *Wikipedia.org*.

Interpreting a seemingly unlikely phenotypic or behavioral outcome in such a context can thus allow to explain it. Gazelle stotting (**Figure 3**) for instance be understood as an honest signal of its ability to outrun an incoming predator: the higher it jumps, the longer it can be expected to evade a predator – and it is strategically optimal for predators to decide which prey to attack (if any) based on how high they jump. With respect to bird plumage,

⁶If one assumes that signaling level (sender strategy) and inferring quality from signals (receiver strategy) are biologically encoded and inheritable, and therefore subject to natural selection – which converges to ESS equilibria.

blue plumage in male grosbeaks at a female audience of potential mates, although it remains unclear which specific quality is signaled – leaving room for debate. Finding the correct underlying quality associated to a (potential) signal, as well as its audience, is not usually straightforward, as decisions such as sexual partner choices involve a variety of overlapping elements and decisions processing shaped by natural selection – potentially even including several qualities and corresponding signals

1.3.2 Social signaling and human prosocial behavior

The theory of costly signaling has also been used by researchers to help shape our understanding of human behavior, starting with economists. Over a century ago, thus framed higher-class luxury consumption and leisure as signals of their status and/or wealth. The underlying logic and assumptions of economic signaling theory is the same as before (Spence, 1974), with honesty emerging from competition⁷ (and differential costs), providing economists with a framework for understanding costly behavior that is not immediately followed by material benefits.

Human prosocial behavior and its underlying motivations can be framed in this light. By definition, prosocial motivations push us to act towards the benefit of others, thus often paying a cost (time, money...) that is not immediately followed by benefits of the same nature (as in **Figure 2**). However, prosocial behavior may entail benefits of a social nature along the road, which can be captured by the concept of reputation behaviors and their underlying prosocial motivations, such as equitable sharing and equity essential building-blocks in our understanding of the emergence of reputation and prosocial motivation in a specific context

In the context of the internship, prosocial motivation was approached using costly social signals, whose purpose are to broadcast qualities which serve as bases for the establishment of social relations. Social signals are evolutionarily stable if they are correlated with qualities which increase the fitness of members of the audience. For instance, if in a certain context such as intergroup conflict, being acquainted with brave individuals increases one's fitness, then such individuals will be in demand, and displaying signals correlated with courage becomes a valid strategy, up to a certain cost.

⁷Which is a fundamental element of Darwin's Competition between signalers pushes them to signal at higher levels, up to the point where marginal gains and marginal costs cancel each other out, making their signals honest indications of their inner payoff structure – which is determined by their quality.

1.3.3 Social signaling and high-cost prosocial behavior

Signaling competition can have non-linear effects, as individual decisions depend on decisions taken by all or several individuals, leading to the emergence of behavior at the collective level. Thus, unconstrained competition relations typically leads, at the collective level, to the emergence of a signaling “elite”, which captures the lion’s share of asymmetrical social affiliations, and a “silent” majority (see **Figure 4**). In such a context, individuals who are unable to attain the levels of the elite have nothing to gain from signaling and should theoretically refrain from doing so⁸. Real-world examples include the emergence of “saints” in rural Morocco or the emergence of individuals with thousands if not millions of followers inside each Twitter community

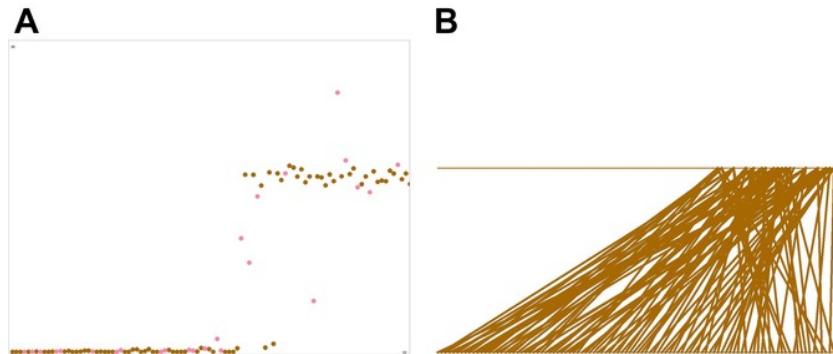


Figure 4: Emergence of an “elite” in the case of unconstrained competition (Dessalles, 2014; figures are his). Individual (dots) signal levels are represented on the left (**A**), and (asymmetrical) social bonds on the right (**B**).

Competition for signaling can therefore lead to extreme costs. In some situations, bravery may even be signaled by strong risk-seeking behavior – so long as the associated costs (significantly higher death probability) are upset by even higher benefits (related to the social advantages of being a member of the “elite” in this situation). Thus, while bravery may be supported by genuine prosocial motivations (a genuine concern with the benefit of the group, to the point of risking one’s life for it), prosociality and collective benefit should be understood as part of (signaling) bravery, not its explanation (Patton,

⁸Somewhat similarly, gazelles or other prey who are unable to jump high enough (due to injury, their age...) should refrain from stotting, leading to segregation between signalers and non-signalers. Collective dynamics matter only however to the extent that stotting may depend on the behavior of other gazelles (as being the least fit gazelle in a group should also lead to not signaling), i. e. to the extent to which the audience should be understood as other gazelles, not the predator.

1996). In contrast, when social relations are constrained to be symmetrical (as on **Figure 5**), competition typically leads to generalized signaling, with individuals pairing up horizontally with individuals of similar quality. In such an example, one would expect benefits, and therefore signaling cost, to remain manageable. Real-world examples may include friendship, where the constraining symmetrizing factor is time spent together or sexual pairing in grosbeaks and many other bird species, where care for offspring most often involves both parents.

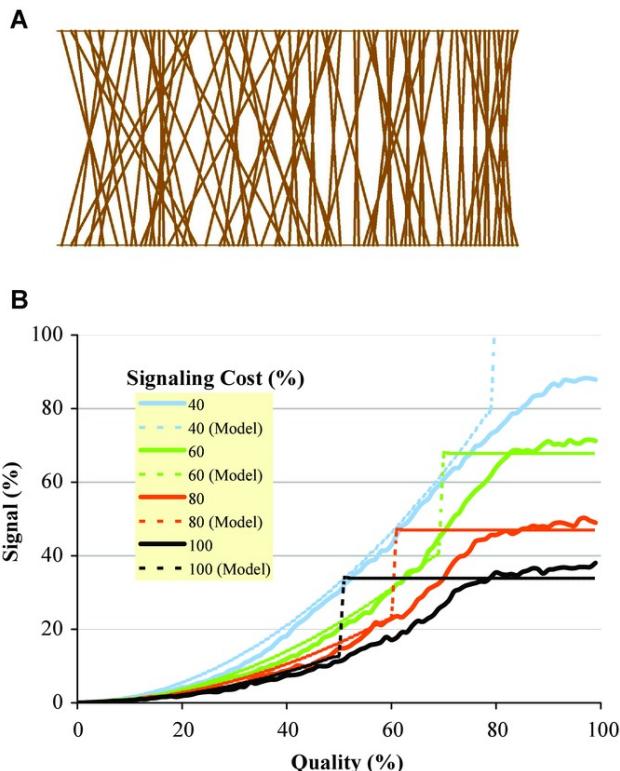


Figure 5: Emergence of generalized signaling (Dessalles, 2014; figures are his). Signaling value as function of individual quality is represented beneath (B), for differing costs; (symmetrical) social bonds above (A).

1.3.4 Self-sacrifice: a two-tier social signaling model

In a context akin to intergroup conflict, one can imagine that commitment to the group would also be a socially in-demand quality. Betrayal by friends may for instance be a real possibility: in such a context, choosing friends who will not betray you for the enemy would be crucial (and correlate with your

fitness), paving the way for costly signals intended to demonstrate such commitment, associated with achieving higher social status. Thus, hypothesis (**H2**) paves the way for addressing question (**Q2**) within a signaling model:

Hypothesis 2 (H2): *Betrayal by friends is an important feature of inter-group conflict. When conflict between groups rises in intensity, it may become a very real possibility (a probable event), carrying dire consequences for the betrayed individual (significant cost). In such a situation, individuals may legitimately fear being betrayed by friends who are disloyal to the group (who may betray in favor of the over warring party).*

This explanation cannot hold for self-sacrifice as defined here, as death has been assumed to be certain. Costs associated with self-sacrifice are too large; signalers would not survive to enjoy the advantages of their new status. Another hypothesis was thus added to the model studied during the internship: that *social status be in part heritable for our species*, as the high status of an individual can raise that of every member of his or her family. More specifically, in a situation where self-sacrifice can be understood as a signal of one's commitment to the group, with high potential benefits which senders never enjoy, it is hypothesized that these social benefits should (at least in part) *spill over* to members of their family. Hypothesis (**H3**) thus paves the way for addressing question (**Q4**), provided that martyrs are indeed revered and memorialized⁹ (or "honored"; see under).

Hypothesis 3 (H3): *Status is inheritable in our species*¹⁰.

Seeing the extreme nature of the costs envisioned, self-sacrifice can only be stable if it remains a low-frequency behavior (the fewer the heroes, the higher their status), pushing us to relate this hypothetical signal with unconstrained competition at a *global* level.

At a more *local* level, individuals have an incentive to advertise their commitment to the group in a less costly way, in order to form more symmetrical friendship bonds, as explained above. The idea is that for such individuals, the self-sacrifice of a group "hero" may constitute an opportunity: if such behavior is understood globally as demonstrative of commitment to the group,

⁹The group under study is implicitly assumed to retain some form of "memory" of its martyrs. **Section 3.1.2** will make this a bit more explicit.

¹⁰This hypothesis could be relaxed if one assumes that honoring focuses on an element shared by "heroes" and their relatives – such as their last name. This seemed like unnecessary hair-splitting however, as the fact that names are shared in a family (and bear status) is probably itself a reflection of the heritability of social status in our species.

then costly behavior associated to a hero's self-sacrifice may also be interpreted as such by potential friends. Thus, at a local level, individuals may signal their commitment to the group by *honoring* such fallen heroes, for instance by engaging in conspicuous ceremonies in their name, in order to attract friends.

If and when honoring emerges as an honest signal of one's commitment to the group, emitted by all who seek friends, this may create the conditions necessary for self-sacrifice itself to emerge. In a situation of generalized honoring, the (artificial) social benefits associated with being a (rare) hero, which spill over to relatives (translate into actual social benefits), may be enough to compensate the extreme costs of self-sacrifice. Self-sacrifice may then theoretically be beneficial to an individual's inclusive fitness (see **Section 1.2.2**).

The conjecture investigated during this internship, in response to its research question is thus:

Conjecture: *In a biological population,*

- *Engaged in intergroup conflict, where betrayal by friends is a real possibility (**H2**) ;*
- *For which social status is heritable (**H3**) ;*

Self-sacrifice and honoring may emerge as first-order and second-order signals of individuals' commitment to the group.

1.3.5 General predictions

The explanation investigated during this internship thus involves a two-tier signaling model. Self-sacrifice and honoring are envisioned as first-order and second-order signals of commitment to the group, which cannot exist without one another, and which take on such meaning because of one another (honoring is intended to signal commitment by referring to self-sacrifice, thus making self-sacrifice a first-order signal of commitment).

Both patterns are expected to be *mutually reinforcing* (**P2**). With respect to first-order signaling, potential individual benefits grow with the number of individuals engaged in second-order signaling (and their average level of signaling) – although, as mentioned above, self-sacrifice should remain a low-frequency behavior (**P1**). Conversely, the higher the visibility of a first-order signal, the more one stands to gain by emitting a second-order signal based on it.

Prediction 1 (P1): *If and when self-sacrifice emerges, it remains a low-frequency behavior.*

Prediction 2 (P2): *Honoring and self-sacrifice are mutually reinforcing.*

2 | Methodes used during the internship

2.1 Objective and methods

2.1.1 Objective

The general objective of the internship is thus to study an alternative explanation for self-sacrifice, using a social signaling framework – as introduced above. Evolutionary signaling theory offers a path for avoiding the potential pitfalls (see **Section 1.2**) of maladaptation and group benefit, as the previous hypothesis was formulated directly and purely in terms of individual (inclusive) fitness.

The underlying objective of such an evolutionary approach to self-sacrifice is to go beyond its various manifestations and/or proximate causes, in order to provide the bases for developing a more integrated understanding of the phenomenon. By investigating these characteristics in relation to a potential biological function of self-sacrifice, one hopes to eventually be able to paint a picture of the causal relations between its various features, and tackle the question of its history

Of course, the actual objectives of the internship are not that lofty. The point is to attempt to provide a logical and reasonably robust argument (see next section), which may allow for limited reinterpretation of self-sacrifice. To begin, an evolutionary account in a general context (intergroup conflict) constitute an argument against the explanatory power of specific beliefs or ideology. Further, if self-sacrifice may be interpreted as a social signal of one's commitment to the group, then group benefit should be understood as part of the signal, not its explanation. In such an interpretation, the fact that a group may benefit from an individual's self-sacrifice is merely a

consequence of that individual engaging in a behavior which has tended to augment his/her inclusive fitness in our evolutionary past.

In addition, the paper aims to provide a stand-alone explanation. As with Blackwell's model, a fundamental idea is that self-sacrifice may be beneficial to inclusive fitness, as a martyr's relatives may benefit from his/her celebrity. In contrast with Blackwell however, the above conjecture aims to provide a mechanism by which may arise such a situation where martyr's gain posthumous celebrity and this entails social (rather than material) benefits to their kin (thus attempting to address **Q3** and **Q4**).

2.1.2 Methods

Work during the internship relied first on computer simulations, using evolutionary paradigms. The basic idea behind these simulations is to achieve a *proof of concept*, meaning the simplest possible simulation that allows for the phenomenon under study to emerge in a robust manner (i.e. to be stable for a large array of parameter values). A credible proof of concept is an argument in favor of an explanation's logical and evolutionary plausibility, as pertaining to the objective outlined above – while failing to do so is indicative of important overlooked flaws.

The projected model was implemented into two Python scripts, and, in order to further assess the logic of the model and the validity of its implementation, predictions made on the basis of the above hypothesis were tested (see below).

Internship work was also grounded in analysis in terms of game theory. Using simplifying hypotheses, the objective was again to obtain a proof of concept, in the form of a (simple) mathematical characterization of the phenomena under study.

2.2 Computer simulations

2.2.1 Evolife

Two scripts were written in Python, and set in the Evolife framework – developed by Jean-Louis Dessimales to study various evolutionary phenomena (<https://evolife.telecom-paristech.fr/>). Evolife has been used to study social signals in general and to model language as a particular social signal

Evolife is based on a genetic algorithm. An individual's behavior is controlled by a binary vector (genome). Note that an individual's "genes" are understood here in a wide, evolutionary (or informational) sense¹: they are *schemata*, portions of an individual's genome small enough to survive through evolutionary time and thus be considered as units of selection (

Individuals live, reproduce sexually, and gain points in a (yearly) life game. Two modes of selection can be implemented:

- ranking: individuals are ranked following the points they have obtained, and are granted a number of potential children that is an increasing (non-linear) function of that rank ;
- differential death: individuals are granted life points (related to their yearly points) which protect them from life hazards, thus increasing their life expectancy (and reproductive opportunities).

As detailed below, the proposed scripts exploit this dichotomy in its treatment of the envisioned first-order and second-order signals. Evolife is modular in structure; the script constituted a scenario implemented inside of Evolife. Both scripts can be found in **Appendix A**.

2.2.2 “Exogenous” base script

Before implementing the actual script, a simplified version was implemented (**Exogenous.py**). This script serves as a basis for the full script, introduced below.

In this version, the social value of self-sacrifice is fixed exogenously, and depends on an *Admiration* parameter A . For the model to make sense, we expect that A will control whether or not self-sacrifice emerges and at what level (**SP0**). This model can be seen as a social alternative to the “economic” model developed by Blackwell (2008).

Self-sacrifice is treated as a genetically controlled trait: individuals in the scenario are endowed with a **SelfSacrifice** gene (8 bits) whose relative value corresponds to an individual's probability of engaging in self-sacrificial behavior in a given year. At (potential) equilibrium, we thus expect all individuals to bear similar (equilibrium) probability of self-sacrifice, which should remain small (**SP1**)² so as to correspond to a small yearly frequency of self-sacrifice,

¹This is different from the definition a molecular biologists would use, as it makes no reference to coding for proteins.

²Thus constraining how much increasing A can lead to increasing p .

as per (P1).

Specific predictions:

- *Admiration A controls equilibrium probability of self-sacrifice p (SP0)*
- *Even when $p > 0$ emerges, it remains small (SP1)*

A heroes' sacrifice benefits his/her descendants. Each year, admiration for heroes "spills over" to the rest of the population. An individual's **Share** in this allocation depends on their ascendant's heroism: each time a parent engages in self-sacrificial behavior, his/her children gain 1 share point. Shares are inherited, as individuals are born with the average of their parent's share times a *SacrificeHeredity* parameter h , situated between 0 and 1.

The emergence and stabilization of self-sacrifice is studied following a ranking mode of selection. Admiration points indirectly gained by individuals are converted into **Reproductive points**, by taking the integer part of these points divided by a *ReproGainsThreshold* parameter (typically fixed at 10).

2.2.3 Actual script structure

This script (**Sacrifice.py**) builds on the previous one. The idea is to see if admiration for heroes, and therefore self-sacrificial behavior, can emerge "endogenously", in a context akin to inter-group conflict, as described previously. The previous dynamic and ranking mode of selection is conserved for self-sacrifice, while other genes (see under) follow a differential death mode of selection (as points gained or lost in the social interactions detailed under are later translated into life points).

In addition to the **SelfSacrifice** gene, individuals are born with a **Patriotism** phenotype, which represents their commitment to the group (quality). For technical reasons, **Patriotism** is either equal to 1 or 0: individuals can either be patriots (1) or non-patriots (0) – who may betray their friends (see **Section 2.2.4**). This simplification is due to the fact that the simulation implements genetic "learning": while a continuous representation of patriotism (and propensity to betray) would have been preferable, this would have required infinite number of genes (one per possible eventuality of honoring – see under).

Hypothesis 4 (H4): *The population under study can be understood as composed of two types of individuals: patriots and non-patriots.*

(H4) is however a powerful guarantor of the simulation's simplicity. To the extent that such a model can produce results that are not artefacts of its binary nature, it may produce an easily understandable proof of concept.

Individuals are also endowed with a **Demand** gene, as well as a **Patriot** and a **NonPatriot** gene, which respectively control their yearly investment in honoring heroes when they are patriots or not.

When there are heroes to honor, this (costly) investment becomes a visible signal of their commitment to the group (patriotism). Individuals select their friends according to the potential signalers they encounter (if there are no heroes, every individual signals at 0) and their **Demand**: any individual who signals above this value may be accepted as a friend (one interaction yields one new friend at the maximum).

Friendship is assumed to be mutually beneficial: an individual with a low **Demand** will thus increase his/her chances of gaining from friendship. However, some **NonPatriot** individuals may be traitors: befriending them ends up being extremely costly, as they betray their friends in the final stage of the year, to their own gain. Thus, in uncertain conditions where being betrayed is a probable and costly outcome, individuals with high **Demand** should fare better.

Expected cost of betrayal is thus expected to be a decisive factor, which should control the emergence of second-order signaling, as captured by **Patriot** and **NonPatriot (SP'0)**. More precisely, we expect the resultant signal to be "honest" – in that, provided expected cost of betrayal is high enough, patriots should end up signaling at significantly higher levels than non-patriots, thus guaranteeing that the signal remains *informative* to its audience (**SP'4**).

Following (**P2**), we expect second-order signal levels – and particularly those of patriots, following (**SP'4**) – to in turn control the emergence of **SelfSacrifice**, with probability p remaining small following **P1**.

Specific predictions:

- *Probability and cost of betrayal (and their product, expected cost of betrayal) control the emergence of second-order signaling (SP'0)*
- *Even when $p > 0$ emerges, it remains small (SP1)*
- *First-order and second-order signals are mutually reinforcing. Thus, expected cost of betrayal should also control the emergence, and to the extent that p remains small, the level of self-sacrifice (SP2)*

- When both signals co-emerge, honoring should be honest to some extent – so that individuals may use it to determine signalers' quality (*SP'4*)

2.2.4 Patriots and non-patriots

The population under study is thus made up of only two types, patriots and non-patriots. Seeing the centrality of betrayal in our proposed explanations, these qualities are tied to the possibility of betrayal in the simplest manner: non-patriots may betray their friends, while patriots are assured not to. Arbitrarily, these types each make up 50% of the population, a proportion which cannot be modified in the simulation.

Number of traitors however is configurable: when parameter *NbTraitors* is equal to t , there are thus $t/2$ traitors in total in the population. When betrayal occurs, the aggrieved party bears a cost of *DenunciationCost* (typically 100), while the 'traitor' is rewarded *Judas* (25). Otherwise, friendship entails a benefit of *FriendshipValue* (10). In any case, 'joining forces' with friends is expected to be beneficial – perhaps particularly in time of uncertainty, being isolated is detrimental. Individuals obtain a linear bonus of *JoiningBonus* (10) times their number of friends.

The simulation implements two 'modes', which puts a relative strain on non-patriots second-order signaling. In the first, non-patriots honoring signals are capped by *MaxOffer* (typically 50 out of 100).

The idea is that these potential traitors should already have other investments – in the interpretation whereby they betray in favor of the other warring group,

(in the rival group, in) Non-patriot's offering is capped: they already have other investments (in the rival group, in themselves...) If social demand is too high, this could push them (perhaps similar to 'exclusive' signals: one can't wear clothes tied to two social classes at once, or bear tattoos from two rival gangs...)

Honoring is more costly for non-patriots: investing time (for instance) in the group is less-interesting than for patriots, since they aren't as committed to the group and thus stand to gain less in relative terms purely in relation to the group (opportunity cost)

As with any signaling model, honesty at equilibrium is key. While this does not appear to be a problem with respect to self-sacrifice, seeing its irremediable nature, explaining why honoring should be an honest signal is crucial.

The idea is that the maximum amount which a dishonest signaler can invest in honoring should be capped to a lower level than that of an honest signaler, since he/she does not intend to be faithful to the group in the long-run and thus stands to gain less in terms of social benefits directly tied to such a group. Another way of looking at this is to consider that a dishonest signaler should face a marginal cost premium, since investing time in the group is less interesting, as the signaler stands to gain relatively more from other sources (other groups, his- or herself).

this could be the case if non-patriots honoring potential is capped ('MaxOffer' parameter) or if they face a signaling premium ('DishonestPremium').

- that when 'NbTraitors' is high enough for there to be a selective pressure for having **Demand** above 'MaxOffer', honoring is a potentially honest signal. If, in addition, the benefits to patriots outweigh the costs, then honoring may emerge at such a level. Seeing as we chose 'MaxOffer' at a level which, in the "exogenous" script (for 'Admiration') was sufficient for self-sacrifice to emerge and stabilize, we that, provided that honoring can emerge, that self-sacrifice emerges as well ; We expect 'MaxOffer' to play a decisive role, as it should correspond to the potential equilibrium state – when it is attainable. Testing 'MaxOffer' at 100 would be a way of seeing the consequences of dishonest signaling: under normal conditions we would expect that **Demand** then rise to the maximum (as it is impossible to select patriot friends), precluding any signaling from emerging. However, if 'Judas' is high enough, signaling may paradoxically remain interesting for non-patriots and we would expect a dishonest signaling equilibrium to emerge.

In extreme situations, competition can thus lead to large individual costs, hence honesty (see **Section 1.3.2**). Correlation between adversarial conditions and cost of signaling commitment has been observed in military and paramilitary groups (tattoos) may emerge This can allow enable a re-framing of the role played by painful rituals in relation to identity fusion and self-sacrifice. As argued in the previous section, competition ensures the honesty of signals of one's commitment to the group, whose cost grows with risk to the group (with demand). One is tempted to interpret painful rituals in this light, which allows to account for the correlation between their intensity and that of fusion, as felt by the group's members Thus, "rites of terror" example of a social signal, which may emerge when groups face extraordinary threat. In addition, the proposed two-tier model offers a way to connect fusion and such rites with self-sacrifice, tentatively³ suggesting that in such conditions,

³This is merely a suggestion, one could probably play this argument the other way around.

demand may allow extremely costly rights to emerge, which may warrant self-sacrifice – as mediated by the proximate feeling of “extreme” identity fusion.

2.2.5 Main variable parameters

With respect to signal honesty: - ‘DifferentialCosts’: specifies the “mode” – equal to 1 when non-patriots face a premium, and 0 when they face a cap (0 by default); - ‘DishonestPremium’: premium (percentage) faced by dishonest signalers in differential costs mode; - ‘MaxOffer’: maximum investment that non-patriots can afford in honoring (out of 100). Fixed at 50 in an initial stage, although exploring other values – including 100 – will be useful (see after). With respect to the social context, including the risk due to intergroup conflict (betrayal): - ‘NbTraitors’: probability that a non-patriot actually ends up being a traitor. This is the main parameter, which will vary between 0 and 100; - ‘FriendshipValue’: value gained from being friends with a non-traitor – fixed at 10 in an early stage; - ‘JoiningBonus’: value gained from having been accepted as friend by someone else – fixed at 10 in an early stage; - ‘Judas’: what a traitor gains from betraying you – fixed at 20 in an early stage; - ‘DenunciationCost’: cost of being betrayed – fixed at 100 in an early stage. Individuals start off the year with 100 points. The previous values for social points (last four parameters) were obtained via reasonable estimates from this scale, and after numeric simulations of honoring in isolation (see Honor.py). Given all these parameters, ‘NbTraitors’ controls the probability of betrayal. In order to differentiate with the cost of betrayal, simulations explored different values for the latter two parameters. In this simplified version with only two levels of patriotism, ‘MaxOffer’ creates an arbitrary cutoff, above which a signal can be considered as honest. For this reason, ‘MaxOffer’ will be largely kept constant at the arbitrary 50,

2.2.6 Other fixed parameters (at least in a first stage)

With respect to the (expected) first-order signal: - ‘ReproGainsThreshold’: corresponds to the value in terms of points acquired through honoring of one’s descendants that yields an additional reproductive unit (for ranking) – kept at 10; - ‘Selectivity’: controls how the degree of selection by ranking (how much expected number of offspring increases with rank); - ‘SacrificeHeredity’: controls how shares are inherited from parents – fixed at 50 after an early analysis of results in the “exogenous” case; - ‘ReproductionRate’: expected

rate of children left each year. Fixed at 15

With respect to the (expected) second-order signal: - ‘HonoringCost’: percent of investment in honoring that translates into cost – fixed at 100; - ‘SelectionPressure’: maximum number of life-points that can be earned ‘in a round – fixed at 6 (meaning that individuals with the highest scores have to be randomly selected 7 times before enough other individuals are selected between 1 and 6 times, in order to die that turn); - ‘EraseNetwork’: indicates whether individuals’ friendship networks are reinitialized each year. Fixed at 0 (False), mainly in order to avoid having to largely increase ‘Rounds’ and ‘NbInteractions’ (and hence computation time). - ‘Rounds’: number of times friendship-forming interactions are launched in the population – fixed at 10; - ‘NbInteractions’: number of such interactions per round, where one randomly chosen individual may accept a maximum of one friend – fixed at 100; - ‘SampleSize’: size of the random sample of (live) people an individual thus interacts with – fixed at 5; - ‘MaxFriends’: maximum number of friends one can hold – fixed at 10. Friendship bonds are assumed to be symmetrical here;

Genetic and general population parameters: -‘AgeMax’: age after which individuals automatically die – fixed at 40; - ‘PopulationSize’: fixed at 20013; - ‘MutationRate’: fixed at 5 per 1000; - ‘NbCrossover’: number of crossovers occurring during sexual reproduction – fixed at 1; - ‘GeneLength’: fixed at 8, weighted bits (meaning that each gene value corresponds to a value between 0 and 2 to the power of 8 minus 1); - ‘AgeAdult’: age before which reproduction is impossible – fixed at 0.

2.3 Predictions

Specific predictions for these simulations are derived from those made in **Section 1.3.5**). The hypothesis under study translates into: under plausible conditions corresponding to the default parameter values detailed above (representing a context of intergroup conflict), honoring and self-sacrifice should co-emerge as second and first-order signals of patriotism (inability to betray). Both of these signals, as captured by individuals’ equilibrium gene values, should be mutually reinforcing. At equilibrium, neither should remain stable in the absence of the other, and large levels of one should correspond to large levels of the other. When and if self-sacrifice emerges, however, it should remain a low-frequency behavior, captured here by low probability of self-sacrifice for all individuals at equilibrium, as controlled by relative gene

value.

More specifically, keeping all other parameter values constant, expectations are: - that for ‘NbTraitors’ equal to 0 or sufficiently small, **Demand** should stay at 0 or non-significantly higher (as friendship is risk-free), thus inducing **Patriot** and **NonPatriot** to stay low, preventing the emergence of self-sacrificial behavior; - that when ‘NbTraitors’ is high enough for there to be a selective pressure for having **Demand** above ‘MaxOffer’, honoring is a potentially honest signal. If, in addition, the benefits to patriots outweigh the costs, then honoring may emerge at such a level. Seeing as we chose ‘MaxOffer’ at a level which, in the “exogenous” script (for ‘Admiration’) was sufficient for self-sacrifice to emerge and stabilize, we that, provided that honoring can emerge, that self-sacrifice emerges as well ; - if we vary ‘DenunciationCost’ instead of ‘NbTraitors’, we expect similar dynamics ; - if we decrease ‘JoiningBonus’ under a certain threshold, attracting friends through signaling patriotism should become overly costly (risky), preventing any non-null behavior at equilibrium.

We expect ‘MaxOffer’ to play a decisive role, as it should correspond to the potential equilibrium state – when it is attainable. Testing ‘MaxOffer’ at 100 would be a way of seeing the consequences of dishonest signaling: under normal conditions we would expect that **Demand** then rise to the maximum (as it is impossible to select patriot friends), precluding any signaling from emerging. However, if ‘Judas’ is high enough, signaling may paradoxically remain interesting for non-patriots and we would expect a dishonest signaling equilibrium to emerge.

3 | Results

3.1 Output

3.1.1 Typical simulations

A simulation = show picture : genes, field, network => show two picture, one with also model SS2 Genes = weighted over 8 bits : values between 0 and $2^8 - 1$

this should be much longer

so as figures don't run into one another...

otherwise learn to skip pages...

Mostly will be working with average for relative value for genes over the pop / after a lot of time One experiment => results for four genes Typically (unless specified otherwise), values averaged over 30 experiments

3.1.2 Remembered Heroes

These output values are precise to the percentage point. For **SelfSacrifice** this poses a problem, as final values are expected to be relatively small - with typical parameter values (as in **section 2**), average probability of self-sacrifice is capped at 2-3%.

For this reason, simulations also kept track of **RememberedHeroes**, a measure for the number of heroes in the live memory of a society. At each simulation step (year), the number of "voluntary" martyrs was computed by subtracting the expected number of martyrs which could be solely attributable to mutations in a population that did not engage in self-sacrifice. These were added to the number of heroes which could be assumed to have

been *witnessed* by (alive) individuals in the population. In a given year, **RememberedHeroes** corresponds to the number of heroes witnessed by at least *RemThreshold* percent of the population (5 % in practice). In situations where self-sacrificial behavior can be said to be absent, **RememberedHeroes** is close to 0 (e.g. when *Admiration* is null in the first model under).

3.2 Exogenous model: simulation outputs

3.2.1 Main results

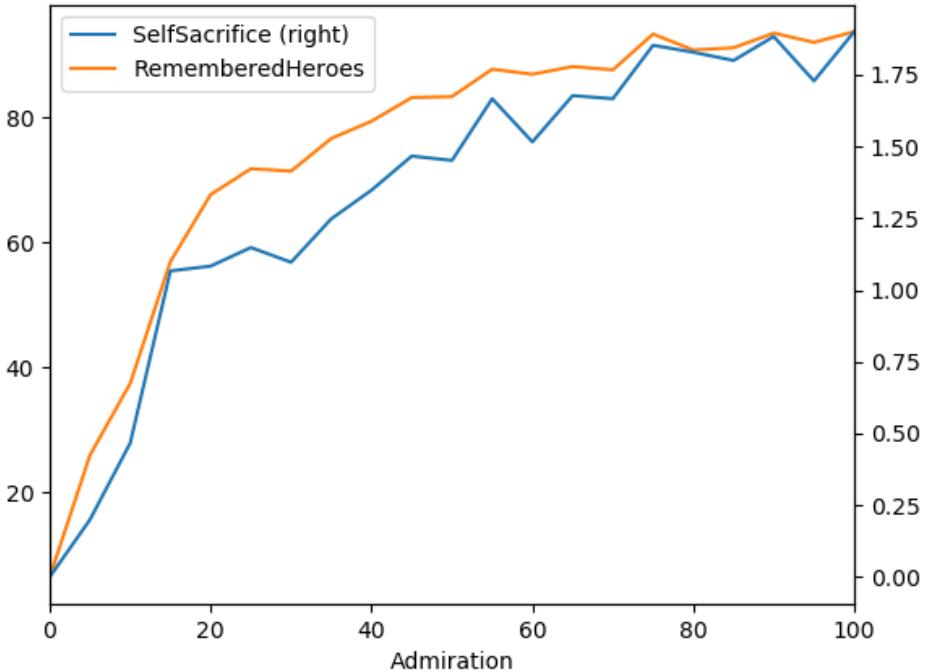


Figure 6: **SelfSacrifice** and **RememberedHeroes**, as a function of *Admiration* (typical parameter values).

Figure 6 shows results obtained by averaging over 30 simulations, according to the *Admiration* parameter (all others being kept constant at the previously described values). As could be expected, **RememberedHeroes** rises from 0 with *Admiration*, quickly reaching two thirds of its maximum value of under 100 heroes when *Admiration* exceeds 20.

SelfSacrifice follows similar dynamics, with values ranging between 0 and 2% (values to a higher precision than the percentage point being obtained artificially by averaging over experimental results). These values are far from negligible: for a typical population of 200, we expect an average of (almost) 4 martyrs each year. Such collective behavior is captured here by individuals all bearing similar genetic probability P of self-sacrifice at equilibrium (and obtaining similar **Reproductivepoints** in the long-term), as seen on **Figure 7**.

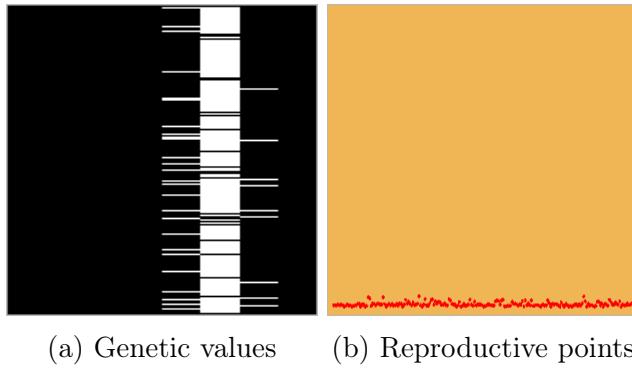


Figure 7: Snapshot of individual values, $A=50$. An individual's genome is represented by a horizontal line on the left: here most bear a **SelfSacrifice** gene of relative value $\frac{2^2}{2^8-1}$ (around 1.6%). Individuals (horizontal axis) and their reproductive points (vertical) are represented on the right.

However, one can also imagine the collectively equivalent situation where only a small fraction f of individuals engage in self-sacrificial behavior with higher probability p (with $P = f * p$), to the much more significant benefit of their families (see **Section 3.3**).

3.2.2 Influence of *ReproGainsThreshold RGT*

As a function of *Admiration*, **RememberedHeroes** resembles a function of the form: $f_{C,\tau}(t) = C * (1 - \exp(-t/\tau))$. Very approximately¹, it can best be approached by a function of this form with parameters $C = 95$ and $\tau = 20$, as visible on **Figure 8**.

¹By choosing optimal C and τ at a precision of 5 units; the idea here being simply to get a "feel" for overall variation with A . Optimal parameters are the ones for which Euclidean distance is minimal.

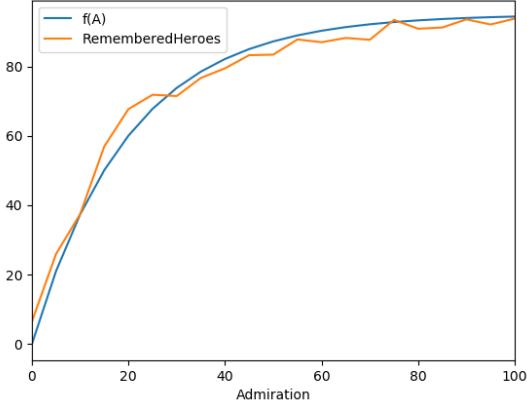
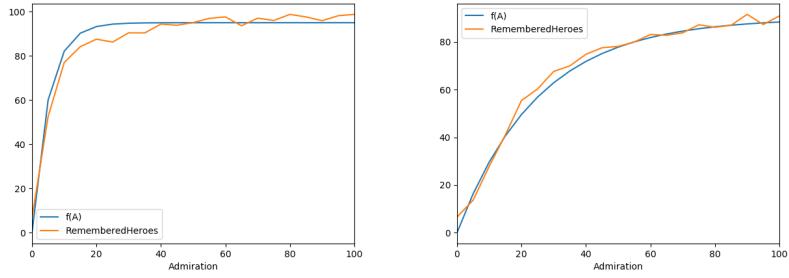


Figure 8: **RememberedHeroes** and corresponding optimal f as functions of *Admiration*, for $RGT = 10$ ($C = 95, \tau = 20$).



(a) $RGT = 5$ ($C = 95, \tau = 5$) (b) $RGT = 20$ ($C = 90, \tau = 25$)

Figure 9: **RememberedHeroes** and f for $RGT = 5, 20$

The key parameter governing relative growth of **RememberedHeroes**² thus appears to be one equal to two times *ReproGainsThreshold*. The importance of RGT is not surprising since it plays the crucial role of defining the unit in which **Reproductivepoints** are counted. The relationship between RGT and final results is however non-trivial, since variation according to A for $RGT = 5$ (resp. $RGT = 20$) are best captured by $\tau = 5$ (resp. $\tau = 25$), as visible on **Figure 9**; suggesting perhaps a quadratic relationship between RGT and τ . **Section 3.3.2** returns to this issue.

²This output's absolute value is arbitrary, chosen according to *RemThreshold* in order to provide for visible variations (see **Section 3.1.2**).

3.2.3 Variation with *SacrificeHeredity* h

Contrary to what could be expected, $h > 1$ is a problem: probs because pop = too small and all end up with same h... - as shown by simu // because same family ? (or just expected in LT?) $h = 0$ discontinuity = expected

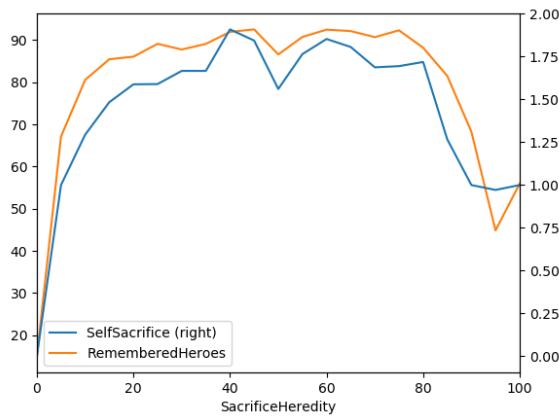


Figure 10

3.2.4 Variation with *ReproductionRate* r

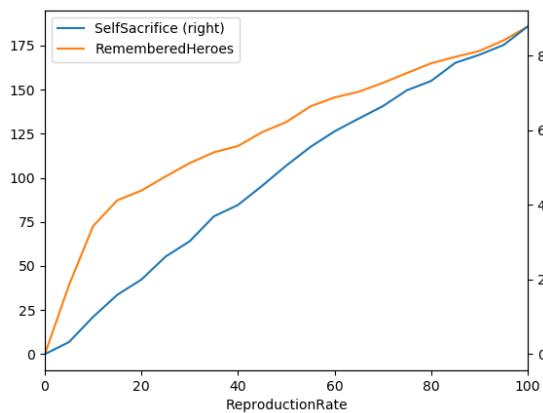


Figure 11

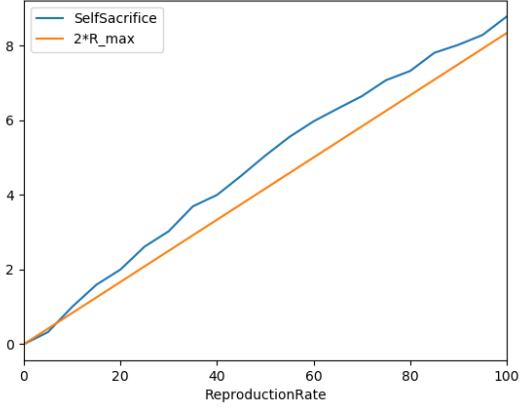


Figure 12

3.3 Exogenous model: mathematical proof of concept

3.3.1 Simplifying assumptions and characterization of equilibrium

Individuals live to a maximum of *AgeMax* M years, fixed at 40 here. Expected life span is lower however, as individuals face random accidents (in a wide sense) and is equal to $\alpha * M$, where α captures the effects of natural selection. In this model, all individuals face the same α , but may obtain differing reproductive opportunities (**Selectivity** mode).

An individual who engages in self-sacrificial behavior with probability p shortens his/her expected lifespan by a multiplicative factor of (see **Appendix B.1.2**):

$$\beta(p) = \frac{(1 - p) - (1 - p)^{(M+1)}}{p * M} \quad (1)$$

As such, said individual's reproductive window is shorter, which should lead to the disappearance of self-sacrificial behavior - unless this is compensated by increased reproductive opportunities. In contrast to how the actual simulation is played, let us assume that:

- Only a *negligible proportion* f of individuals engage in self-sacrificial

behavior, with equal probability p ;

- These future "heroes" will be the first this society sees (everyone starts off with no **Reproductivepoints** RP);
- Generations do not overlap;
- Individuals are granted their lifetime reproductive potential R at birth (in contrast with a year-by-year attribution).

In such a situation, individuals who engage in self-sacrificial behavior obtain on average $\alpha * \beta(p) * M * r$ offspring, with others obtaining $\alpha * M * r$, where r is the population-level *ReproductionRate*. However, the former's children are granted a larger reproduction potential $R_+(A, f)$, which notably depends on *Admiration A*. Since f is assumed to be very small, $R_+(A, f)$ is the same for all children of individuals who self-sacrifice (we neglect the possibility of individuals being born from two would-be martyrs), and the children of non-would-be martyrs obtain a reproductive potential which can be approximated to r .

If, in addition, we assume that p and M are sufficiently large³ for would-be heroes to largely end up actually laying their life for the group (and not dying in another way), such individuals obtain on average $R_+(A, f) * \alpha * \beta(p) * M * r$ grand-children, while others obtain $\alpha * M * r^2$. Thus, a first-order⁴ (neglecting the effects on subsequent generations) characterization of equivalence between both strategies can be written as:

$$R_+(A, f) * \beta(p) = r \quad (2)$$

3.3.2 Necessary condition for an ESS

Let N be equal to *PopulationSize*, implicitly assumed to be large here (since f is negligible).

In the spirit of the simplifying assumptions made above, let us assume that total social admiration is borne the children of non-heroes, and is thus equal to $A * (1 - f)N * \alpha Mr$. For children of martyrs, two cases are possible:

³We still expect $P = f * p$ to be relatively small, to avoid "martyr over-crowding", but for this to be primarily due to f , in the current mathematical characterization.

⁴Note that apart from (3) equations in this section are valid at best at the first-order in $f << 1$ and are computed for expected values.

- either $\frac{A*(1-f)N*\alpha Mr}{fN*\alpha\beta(p)Mr} < RGT$, and all receive reproductive potential r ;
- or $\frac{A*(1-f)N*\alpha Mr}{fN*\alpha\beta(p)Mr} \geq RGT$, and all receive $R_+(A, f) > r$.

Thus, (p,f) can only constitute an ESS if:

$$A \geq A_{min} = \frac{RGT * f * \beta(p)}{1 - f} \iff f \leq f_{max} = \frac{A}{RGT * \beta(p) + A} \quad (3)$$

3.3.3 Mathematical characterization of the ESS

As shown in **Appendix B.1.3**, a first-order approximation of R_+ in the latter case is (S is equal to *Selectivity*):

$$R_+(A, f) = \frac{S * r}{\log(1 + S)} * \left(1 - \frac{f * \beta(p)}{2}\right) \quad (4)$$

This allows us to deduce first-order approximations for $\beta(p)$ and p at (potential) ESS, as shown in **Appendix B.1.4**:

$$\beta(p) = \frac{r}{R_{max}} + \frac{1}{2} * \left(\frac{r}{R_{max}}\right)^2 * f \quad (5)$$

$$p(S, M, f) = \frac{1}{1 + M * \frac{\log(1+S)}{S}} * \left(1 - \frac{\frac{1}{2} * \left(\frac{\log(1+S)}{S}\right)^2}{1 + M * \frac{\log(1+S)}{S}}\right) * f \quad (6)$$

Using (3), we also deduce that, for such an ESS to exist, we must have:

$$A \geq A_{min}(RGT, S, f) = \frac{RGT * \log(1 + S)}{S} * f \quad (7)$$

3.3.4 Brief Discussion

$p = +$ function of $S =$ duh // - of M : more to sacrifice But **no r**, no $A --- > should be captured by f.... : OK for A in f_{max}... Maybe cheat : admiration goes to heroes = the parents... then to children ? How ? = need to have a generational gap (but cool here that f between 0 and 1)$

- + RGT : here = linear law... probs because of approx $f \ll 1$, all receive the same...
- + h: not here, vu le modele
- + lim for JLD : does not depend on nb of heroes... -> visibility $\log(\text{nbheroes...})$?)

3.4 Two-tier model simulation outputs

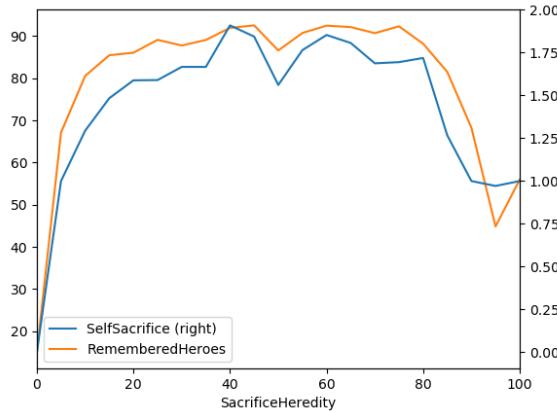


Figure 13

3.5 Two-tier model: mathematical analysis

3.5.1 Self-sacrifice

For would-be martyrs M , the situation can be approached as in **Section 3.3**: the question comes down to whether a non-trivial set of ESS strategies comprising a proportion f of agents engaging in self-sacrifice with probability p can emerge⁵, and meet the requirements of equation (2).

Total social admiration A is now however *endogenous*, as it is determined by

⁵ And $1 - f$ agents not doing so.

how much non-martyrs \overline{M} engage in second-order signaling (honoring h):

$$A = \sum_{h(\overline{M}) \geq V_T} h(\overline{M}) \quad (8)$$

At (potential) equilibrium, A will thus be determined by (potential) equilibrium levels of honoring by patriots P and non-patriots \overline{P} and how high *VisibleThreshold* V_T is set. This, in turn, should determine if a self-sacrifice ESS is possible, and, if so, at what proportion of martyrs f can emerge at equilibrium, as before. Both signals are thus connected via equation (2), which can be tentatively rewritten according to equilibrium levels:

$$R_+(h_P, h_{\overline{P}}, f) * \beta(p) = r \quad (2b)$$

3.5.2 Honoring and social score

This is motivation to investigate the possibility of an honoring equilibrium. Honoring h and demand d serve as bases for establishing friendships, which are mutually beneficial (as controlled by *Friendship Value* F). However, inter-group conflict raises the stakes: friends may be true patriots, to one's benefit (as controlled by *PatriotFriendBonus* P), or not be, leading to the possibility of betrayal with probability t (*NbTraitors*) and at potentially enormous cost DC (*DenunciationCost*).

And patriot individual characterized by d and h , who obtains a number of friends k on the bases of these (genetic) characteristics will see his social score vary according to:

$$\mathbf{E}(\Delta_P(d, h)) = k * F + k_P * P - k_{\overline{P}} * t * DC \quad (9)$$

In contrast, a non-patriot individuals also stand to gain *Judas* J if and when they betray their friends:

$$\mathbf{E}(\Delta_{\overline{P}}(d, h)) = k * F + k_P * P - k_{\overline{P}} * t * DC + t * J \quad (\overline{9})$$

3.5.3 Equilibrium when $t \leq \frac{F}{FC}$

Let us suppose that individuals of the same quality all signal at the same level s_P or $s_{\overline{P}}$, and have equal demand d_P or $d_{\overline{P}}$ (noted d when there is no ambiguity).

Since individuals encounter patriots and non-patriots with equal probability, and simply chooses whether or not to accept them based on d (potential friends are not ranked), a patriot with demand d can be in one of four situations (where k notably depends on d and time allocated to encountering potential partners):

$\mathbf{E}(\Delta_P(d, h))$	$d \leq s_{\bar{P}}$	$d > s_{\bar{P}}$
$d \leq s_P$	$\frac{k}{2} * (2F + P - t * DC)$	$k * (F + P)$
$d > s_P$	$k * (F - t * DC)$	0

Table 2: Expected social score (patriot)

When t is sufficiently small, **Table 2**'s first three cells are strictly positive: when betrayal is sufficiently improbable, friendship can always be expected to pay off for patriots - as well as for non-patriots, since $\mathbf{E}(\Delta_{\bar{P}}(d, h)) \geq \mathbf{E}(\Delta_P(d, h))$ ⁶. For both, the optimal "strategy" is thus $d = 0$. In addition, since honoring is costly, the best response to $d = 0$ is to not invest in honoring:

$$t \leq \frac{F}{DC} \implies (d = 0, h_P = 0, h_{\bar{P}} = 0) \text{ is the only NE}^7 \quad (10)$$

With typical parameter values, when $NbTraitors$ is smaller than 10%, no second-order signal - and therefore no first-order signal - should emerge.

3.5.4 "Honest" equilibrium when $t \geq \frac{2F+P}{DC}$

When t is sufficiently large, both cells on the left of **Table 2** are negative: patriots cannot afford to have any non-patriot friends. Let us assume, following the first-version of the proposed simulation, that $s_{\bar{P}}$ is bounded by *MaxOffer MO*, and that non-patriots and patriots pay the same cost for signaling.

For patriots, it is then always better to have $d > MO$ than $d \leq MO$. Indeed, in the latter case where all patriots have demand $d < MO$, we are in a domain which is beneficial for non-patriots: they stand to gain more from signaling (by potentially betraying), at no extra cost. Thus, if $d < MO$, we expect $s_{\bar{P}} \geq s_P$ ⁸: the optimal situation represented by **Table 2**'s top right cell is unattainable, and, on average, patriots stand to lose from friendship.

⁶At least when h is small - see after

⁸And $d_{\bar{P}} \leq d_P$

Conversely, if $d > MO$, patriots either obtain no friends or only patriot friends and thus stand to benefit. Any strategy d^+ verifying $d^+ > MO$ thus weakly dominates any other strategy d^- , where $d^- \leq MO$.

Yet, as evoked in introducing equation (9), an individual's number of friends k depends on d . Given two demands above MO , the smaller is always best, since decreasing demand can only lead to increasing number of (patriotic) friends. For patriotic individuals, the optimal strategy is thus the smallest available d which is larger than MO , which we will note $d = MO^+$.

Let us assume in addition that individuals have ample opportunity to meet potential friends, forming Max_F friendships when possible, at cost c , and that $c * MO^+ < Max_F * (F + P)$ ⁹. Patriots stand to gain from signaling at levels above of MO , and for them, the optimal strategy is thus MO^+ , following the same reasoning as above.

Non-patriots, however, can only attract other non-patriots:

$$\mathbf{E}(\Delta_{\bar{P}}(d, h)) = k * (f - t * DC + t * J)$$

$$\mathbf{E}(\Delta_{\bar{P}}(d, h)) \geq 0 \iff t \leq \frac{F}{DC - J}$$

In practice, the latter condition is incompatible with the one presented at the beginning of this section, since, with typical values, $\frac{F}{DC - J} = \frac{2}{15}$ and $\frac{2F+P}{DC} = 0.3$. Thus:

$$t \geq \frac{2F+P}{DC} \implies (d = MO^+, h_P = MO^+, h_{\bar{P}} = 0) \text{ is the only NE} \quad (11)$$

When $NbTraitors$ is greater than 30%, a purely honest equilibrium should thus emerge at the second-order - leading to self-sacrifice at a level corresponding to $A = MO^+/2$, since only half the population honors martyrs.

3.5.5 Dishonest equilibrium when $t \geq \frac{2F+P}{DC}$

In practice not so much: - some dishonest, and not because stabilize at a level below permitted by J : cf ... - RemT = "more optimal" than expected

⁹Otherwise signaling is never beneficial and the only equilibrium is trivially (0,0,0).

$$3.5.6 \quad \frac{F}{DC} < t < \frac{2F+P}{DC}$$

3.6 Second sacrifice model

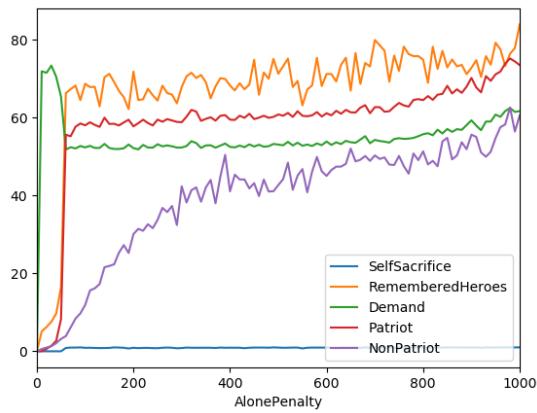


Figure 14

3.7 Limitations

A 'social' version of learning, whereby —> ca apprte quoi ? // ok pas artefact

4 | Perspectives

A | Python scripts

A.1 Exogenous.py

```
1 #####  
2 # EVOLIFE          http://evolife.telecom-paristech.fr  #  
3 # Telecom ParisTech 2018           Jean-Louis Dessalles #  
4 # _____  
5 # License: Creative Commons BY-NC-SA  
6 #####  
7  
8 """ Study of the possibility of self-sacrifice being an ESS  
9    Simplified base version where Admiration is exogenous  
10   Serves as base for Sacrifice.py  
11 """  
12  
13  
14 from random import sample, shuffle, random, choice, randint  
15 from math import exp, log  
16 from time import sleep  
17  
18 import sys  
19 sys.path.append('..')  
20 sys.path.append('../..')  
21  
22 import Evolife.Ecology.Observer as EO  
23 import Evolife.Scenarii.Default_Scenario as ED  
24 import Evolife.Ecology.Individual as EI  
25 import Evolife.Ecology.Alliances as EA  
26 import Evolife.QtGraphics.Evolife_Window as EW  
27 import Evolife.QtGraphics.Evolife_Batch as EB  
28 import Evolife.Ecology.Group as EG  
29 import Evolife.Ecology.Population as EP  
30  
31 from Evolife.Tools.Tools import percent, chances, decrease, error  
32  
33 class Scenario(ED.Default_Scenario):
```

```

34 #####
35 ##### General initializations and visual display #####
36 #####
37
38     def __init__(self):
39         # Parameter values
40         ED.Default_Scenario.__init__(self, CfgFile='Params.evo')
41         # loads parameters from configuration file
42         self.initialization()
43
44     def initialization(self):
45         self.RememberedHeroes = 0 # Nb of heroes that society "remembers" in a given round
46
47     def genemap(self):
48         """ Defines the name of genes and their position on the DNA.
49             """
50         return [( 'SelfSacrifice' )]      # gene length (in bits) is read from configuration
51
52     def display_(self):
53         disp = [( i+1,G.name) for (i,G) in enumerate( self.GeneMap )]
54         L = len(disp)
55         disp += [(L+1, 'RememberedHeroes')]
56         return disp
57
58     def local_display(self, ToBeDisplayed):
59         """ allows to display locally defined values """
60         if ToBeDisplayed == 'RememberedHeroes':
61             return self.RememberedHeroes # displaying a quantity that can be computed within the scenario
62         return None
63
64     def update_positions(self, members, start_location):
65         """ locates individuals on an 2D space """
66
67         # sorting individuals by gene value
68         duplicate = members[:]
69         duplicate.sort(key=lambda x: x.gene_value('SelfSacrifice'))
70
71         for m in enumerate(duplicate):
72             m[1].location = (start_location + m[0], m[1].Reproductive_points)
73             # Allows to place individuals in a 2D field
74
75             ## Other options
76             #m[1].location = (start_location + m[0], m[1].Share )

```

```

75         #m[1].location = (start_location + m[0], m[1].  

76         HeroesRelatedness)  

77         #m[1].location = (start_location + m[0], m[1].  

78     gene_value('SelfSacrifice'))  

79  

80     def remove_agent(self, agent):  

81         """ action to be performed when an agent dies "  

82         # None here, useful for tests  

83  

84 ##### Life_game #####  

85  

86     def life_game(self, members):  

87         """ Defines one year of play (outside of reproduction)  

88             This is where individual's acquire their score  

89         """  

90         # First: make initializations (1)  

91         self.start_game(members)  

92         # Then: play multipartite game, composed of:  

93             # The sacrifices game (2):  

94             AliveMembers = self.sacrifices(members)  

95             # Social interactions between the living (3)  

96             for play in range(self.Parameter('Rounds', Default=1)):  

97                 self.interactions(AliveMembers, nb_interactions=self.  

98                 Parameter('NbInteractions'))  

99                 # Last: work out final tallies (4)  

100                for indiv in AliveMembers:  

101                    self.evaluation(indiv)  

102                    # "Social" scores are translated into life points, which  

103                    # affect individual's survival (not useful here)  

104                    self.lives(AliveMembers)  

105  

106 ##### Life_game ##### (1) Initializations #####  

107  

108     def prepare(self, indiv):  

109         """ Defines what is to be done at the individual level  

110             before interactions  

111                 occur - Used in 'start_game'  

112             """  

113             indiv.Reproductive_points = 0    # Reset reproductive  

114             points to 0  

115  

116     def start_game(self, members):  

117         """ Defines what is to be done at the group level each  

118             year  

119                 before interactions occur - Used in 'life_game'  

120             """

```

```

117     for indiv in members: self.prepare(indiv)
118
119 ##### Life_game ###### (2) Self-sacrifices #####
120 ##### (2a) Individual propensity to self-sacrifice
121
122     ## (2a) Individual propensity to self-sacrifice
123     def deathProbability(self, indiv):
124         """ Converts an individual's genetic propensity to self-
125             sacrifice into a probability
126                 Used in 'selfSacrifice'
127
128         maxvalue = 2 ** self.Parameter('GeneLength') - 1
129         return (indiv.gene_value('SelfSacrifice') / maxvalue)
130
131     def selfSacrifice(self, indiv):
132         """ An agent decides to make the ultimate sacrifice
133             Self-Sacrifice is only a possiblity from a certain
134             age (e.g. adulthood)
135                 and is controlled by the SelfSacrifice gene (see
136             deathProbability)
137                 Used in 'sacrifices'
138
139         p = self.deathProbability(indiv)
140         ## 'Probabilistic' mode:
141         bool = p > random()
142
143         # Other (unused) modes:
144         ## 'Binary mode': (Individuals are programmed to self-
145             sacrifice at a certain age)
146             #bool = p > 0 and (indiv.age > ((percent(self.Parameter(
147                 'SacrificeMaturity')) * self.Parameter('AgeMax'))))
148             ## 'Threshold mode': gene codes for the value of the age
149             after which sacrifice occurs
150             # This is a measure of the 'value' of the sacrifice (how
151             much potential reproduction is given up)
152             #bool = indiv.age > (1-p) * self.Parameter('AgeMax')
153
154             indiv.SelfSacrifice = bool
155             return bool
156
157 ##### (2b) Population-level self-sacrifice game
158     def sacrifices(self, members):
159         """ Self-sacrifice 'game':
160             Heroes may self-sacrifice "for the good of the group"
161             In return they are admired – admiration is exogenous
162             here
163                 Used in 'life_game'
164
165         """

```

```

158     Heroes = []
159     Cowards = members[ :]
160     for indiv in members: # Deciding who are the population's
161         heroes
162             if indiv.SelfSacrifice: # indiv is already a hero (useless here: for earlier version)
163                 Heroes.append(indiv)
164                 Cowards.remove(indiv)
165             elif self.selfSacrifice(indiv): # indiv is not already a hero but is given an opportunity to be one
166                 Heroes.append(indiv)
167                 Cowards.remove(indiv)
168
169             if self.Parameter('Baseline') == 0:
170                 ##### baseline = 0: such that RememberedHeroes is approximately 0 for Admiration = 0
171                 random_heroes = 1 + int( self.Parameter('PopulationSize') * (self.Parameter('MutationRate') / 1000) / self.Parameter('GeneLength'))
172                 self.RememberedHeroes = self.memo_heroes(len(Heroes) - random_heroes, Cowards, threshold = self.Parameter('RemThreshold'))
173                     # Subtracting heroes that can be accounted for by 'pure chance' (mutation)
174             else:
175                 ##### baseline > 0 version (not used)
176                 self.RememberedHeroes = self.memo_heroes(len(Heroes), Cowards, threshold = self.Parameter('RemThreshold'))
177
178                 self.inc_shares(Heroes, Cowards)      # Relatives of heroes gain shares
179
180                 # Heroes are honored (admired) by society
181                 Admiration = self.honoring(Cowards, len(Heroes))
182
183                 # This spills over to their relatives
184                 self.spillover(Cowards, Admiration, threshold = self.Parameter('ReproGainsThreshold'))
185
186                 self.kill_heroes(Heroes)      # Heroes are set to be removed from the simulation
187             return Cowards
188
189     def memo_heroes(self, new_heroes, members, threshold = 5):
190         " Society, through its (alive)_ individuals, remembers its heroes "
191
192         members.sort(key=lambda x: x.HeroesWitnessed, reverse = True)

```

```

192
193     past_heroes = members[ int( percent(threshold) * len(
194         members) ) ].HeroesWitnessed
195         # Heroes pass off to posterity if they are "remembered"
196         by a sufficient amount of alive individuals (threshold %)
197
198     for indiv in members:
199         indiv.HeroesWitnessed += new_heroes
200     return past_heroes + new_heroes
201
202
203     def inc_shares(self , heroes , members):
204         """ Relatives of heroes indirectly benefit from their
205             sacrifice
206             by gaining shares (useful in 'spillover')
207             """
208
209         for Hero in heroes:
210             for Child in Hero.Children:
211                 Child.Share += 1
212
213     def honoring(self , worshippers , nb_heroes):
214         """ Defines the amount of available population-level
215             admiration for heroes
216             (To be overloaded)
217             """
218
219         if nb_heroes == 0: return 0 # No heroes to admire this
220         round
221         return self.Parameter('Admiration') * len(worshippers)
222
223     def spillover(self , members , admiration = 0, threshold = 10):
224         """ Societies' admiration for heroes "spills over" to
225             their relatives
226             who gain reproductive points (Selectivity mode)
227             """
228
229         tot_share = 0
230         for indiv in members:
231             tot_share += indiv.Share
232         if tot_share == 0:
233             return
234         for indiv in members:
235             indiv.Reproductive_points += indiv.Share / tot_share
236             * admiration
237             indiv.Reproductive_points = int(indiv.
238             Reproductive_points /threshold) # Only points above threshold
239             entail reproductive advantage
240
241     def kill_heroes(self , heroes):
242         """ Heroes are programmed to die (see 'lives') """
243         for Hero in heroes:
244             Hero.LifePoints = -1

```

```

232 #####
233 ##### Life_game #### (3) Social interactions #####
234 #####
235 #####
236     def interactions(self, members, nb_interactions = 1):
237         """ Defines how the (alive) population interacts
238             Used in 'life_game'
239         """
240
241         for inter in range(nb_interactions):
242             if not members: return
243             Fan = choice(members)
244             # Fan chooses friends from a sample of Partners
245             Partners = self.partners(Fan, members, int(percent(
246                 self.Parameter('SampleSize') \
247
248                 * (len(members)-1) ))) \
249                 self.interact(Fan, Partners)
250
251     def interact(self, indiv, partners):
252         """ Nothing by default - Used in 'interactions'
253         """
254         pass
255
256     def partners(self, indiv, members, sample_size = 1):
257         """ Decides whom to interact with - Used in 'interactions'
258         """
259
260         #By default, a sample of partners is randomly chosen
261         partners = members[:]
262         partners.remove(indiv)
263         if sample_size > len(partners):
264             return partners
265         if partners != []:
266             return sample(partners, sample_size)
267         else:
268             return None
269
270     def partner(self, indiv, members):
271         """ Decides whom to interact with - Used in 'life_game'
272         """
273
274         # By default, a partner is randomly chosen
275         partners = members[:]
276         partners.remove(indiv)
277         if partners != []:
278             return choice(partners)
279         else:
280             return None
281
282     def interaction(self, indiv, partner):
283         """ Nothing by default - Used in 'life_game' """
284

```

```

278     pass
279
280 ##### Life_game #####
281 ##### Life_game #####
282 ##### Life_game #####
283 def evaluation(self, indiv):
284     """ Computation of social score """
285     # No social score by default
286     pass    # Reproductive points are computed in 'spillover',
287     # (see 'sacrifices')
288
289 def lives(self, members):
290     """ Converts alive members' scores into life points –
291     used in 'life_game'
292     """
293
294
295 ##### Reproduction #####
296 ##### Reproduction #####
297
298 # Reproduction is already defined elsewhere, but some functions
299 # have to be overloaded to create local individuals
300 # This is the case with parenthood (here) and Group.reproduction
301 def parenthood(self, RankedCandidates, Def_Nb_Children):
302     """ Determines the number of children depending on rank
303     Here, Selectivity is non-null by default, leading to
304     rapid (non-linear) convergence
305     (contrary to a purely SelectionPressure mode)
306     """
307     ValidCandidates = RankedCandidates[:]
308     for Candidate in RankedCandidates:
309         if Candidate.SelfSacrifice or Candidate.age < self.
310             Parameter('AgeAdult'):
311                 ValidCandidates.remove(Candidate) # Children and
312                 (dead) heroes cannot reproduce (AgeAdult is null by default)
313                 candidates = [[m,0] for m in ValidCandidates]
314                 # Parenthood is distributed as a function of the rank
315                 # It is the responsibility of the caller to rank members
316                 appropriately
317                 # Note: reproduction_rate has to be doubled, as it takes
318                 two parents to beget a child
319                 for ParentID in enumerate(ValidCandidates):
320                     candidates[ParentID[0]][1] = chances(decrease(
321                         ParentID[0], len(ValidCandidates), self.Parameter('Selectivity'
322                         )), 2 * Def_Nb_Children)
323             return candidates
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850

```

```

317 #####
318 #####
319 #####
320 #####
321 # Adding useful characteristics to individuals
322
323 class Individual(EI.EvolifeIndividual):
324     " Defines what an individual consists of "
325
326     def __init__(self, Scenario, ID=None, Newborn=True):
327         # Specific characteristics pertaining to this scenario
328         self.SelfSacrifice = False # True if individual decides
329         to self-sacrifice this year
330         self.HeroesWitnessed = 0 # Keeps track of heroes that
331         died during individual's life
332         self.Reproductive_points = 0 # Determines number of
333         offspring (Selectivity)
334         self.Share = 0 # Determines share of
335         population-level admiration that spills over to individual,
336         if any
337         self.Children = [] # Keeps track of individual's
338         children
339
340         # General characteristics: individual inherits from the
341         EvolifeIndividual module
342         EI.EvolifeIndividual.__init__(self, Scenario, ID=ID,
343         Newborn=Newborn)
344
345     def inherit_share(self, mom, dad, heredity = 0.5):
346         " Individual inherits shares from his parents "
347         self.Share += (mom.Share + dad.Share) / 2 * heredity
348
349     def isChild(self, Indiv):
350         if Indiv in self.Children: return True
351         return False
352
353     def addChild(self, child):
354         " Adds child to list "
355         self.Children.append(child)
356
357     def removeChild(self, child):
358         """ Removes a child of self
359             Used when the child dies (see 'Group.remove_')
360         """
361
362         if self.isChild(child):
363             self.Children.remove(child)
364
365 #####
366 #####
367 #####

```

```

358 #####
359 #####
360 # Overloading useful elements of Group module with local
361 # Individual class
362
363 class Group(EG.EvolifeGroup):
364     """ The group is a container for individual (By default the
365     population only has one)
366         It is also the level at which reproduction occurs
367         Individuals are stored in self.members
368     """
369
370
371     def __init__(self, Scenario, ID=1, Size=100):
372         EG.EvolifeGroup.__init__(self, Scenario, ID, Size)
373
374     def createIndividual(self, ID=None, Newborn=True):
375         """ Calling local class 'Individual' """
376         Indiv = Individual(self.Scenario, ID=self.free_ID(),
377         Newborn=Newborn)
378         # Individual creation may fail if there is no room left
379         self.Scenario.new_agent(Indiv, None) # Let scenario know
380         that there is a newcomer (unused)
381         return Indiv
382
383     def update_ranks(self, flagRanking=False, display=False):
384         """ Updates various facts about the group
385         """
386
387         # Removing individuals who die of old age (as defined by
388         AgeMax)
389         for m in self.members[:]: # must duplicate the list to
390             avoid looping over a modifying list
391             if m.dead(): self.remove_(self.members.index(m))
392             self.size = len(self.members)
393             if self.size == 0: return 0
394             # ranking individuals
395             if flagRanking:
396                 # ranking individuals in the group according to their
397                 score
398                 self.ranking = self.members[:] # duplicates the
399                 list, not the elements
400                 self.ranking.sort(key=lambda x: x.Reproductive_points
401                 , reverse=True)
402                 if self.ranking != [] and self.ranking[0].
403                 Reproductive_points == 0:
404                     # all scores are zero
405                     shuffle(self.ranking) # not always the same ones
406                     first
407                     self.best_score = self.ranking[0].score()
408             return self.size

```

```

396
397     def update_(self, flagRanking = False, display=False):
398         """ Updates various facts about the group + positions
399         """
400         size = Group.update_ranks(self, flagRanking=flagRanking)
401         if display:
402             if flagRanking: self.Scenario.update_positions(self.
403 ranking, self.location)
404             else: self.Scenario.update_positions(self.members,
405 , self.location)
406             # updating social links
407             for m in self.members: m.checkNetwork(membershipFunction
408 =self.isMember)
409             return size
410
411
412
413
414
415
416     def remove_(self, memberNbr):
417         " An individual is removed from the group and his parents
418         ' family trees "
419         indiv = self.whoIs(memberNbr)
420         for parent in self.members:
421             parent.removeChild(indiv)
422         return EG.EvolifeGroup.remove_(self, memberNbr)
423
424
425
426
427     def reproduction(self):
428         """ Reproduction within the group
429             Uses 'parenthood' (in Scenario) and 'couples' (not
430 reproduced here – from Group module)
431             'couples' returns as many couples as children are to
432 be born
433             """
434             # The probability of parents to beget children depends on
435             their rank within the group
436             # (Except if Selectivity = 0, for which rank becomes
437             random)
438             self.update_(flagRanking=True)    # updates individual
439             ranks
440             for C in self.Scenario.couples(self.ranking):
441                 # Making of the child
442                 child = Individual(self.Scenario, ID=self.free_ID(),
443 Newborn=True)
444                 if child:
445                     child.hybrid(C[0],C[1]) # Child's DNA results
446                     from parents' DNA crossover
447                     child.mutate()
448                     # Child inherits shares: thus grand–children (and
449                     so on) also indirectly benefit from self–sacrifice
450                     child.inherit_share(C[0],C[1], heredity = percent
451 (self.Scenario.Parameter('SacrificeHeredity'))))

```

```

432             child.update() # Computes the value of genes, as
433             DNA is available only now
434             if self.Scenario.new_agent(child, C): # Let
435                 scenario decide something about the newcomer (not used here)
436                     # Child is added to parents' children
437                     lists
438                         C[0].addChild(child)
439                         C[1].addChild(child)
440
441 #####
442 #####
443 #####
444 # Overloading useful elements of Population module with local
445 # Group class
446
447 class Population(EP.EvolifePopulation):
448     """ Defines the population of agents
449         This is the level at which 'life_game' is played
450     """
451
452     def __init__(self, Scenario, Observer):
453         " Creates a population of agents "
454         EP.EvolifePopulation.__init__(self, Scenario, Observer)
455         self.Scenario = Scenario
456
457     def createGroup(self, ID=0, Size=0):
458         " Calling local class 'Group' "
459         return Group(self.Scenario, ID=ID, Size=Size)
460 #####
461 #####
462 #####
463 # Launching of the simulation, using Start module, and which
464 # capabilities to use:
465 # displays over time (average gene value and remembered heroes
466 # here), 2D field, genes, social networks
467
468 def Start(Gbl = None, PopClass = Population, ObsClass = None,
469 Capabilities = 'PCGFn'):
470     " Launch function "
471     if Gbl == None: Gbl = Scenario()
472     if ObsClass == None: Observer = EO.EvolifeObserver(Gbl) #
473         Observer contains statistics
474     Pop = PopClass(Gbl, Observer)
475     BatchMode = Gbl.Parameter('BatchMode')
476
477

```

```

473     if BatchMode:    # Non-graphic mode (for large-scale
474         experiments)
475             EB.Start(Pop.one_year, Observer)
476     else:
477         EW.Start(Pop.one_year, Observer, Capabilities=
478             Capabilities)
479         if not BatchMode: print("Bye.....")
480         sleep(2.1)
481     return
482
483 if __name__ == "__main__":
484     print(__doc__)
485
486 ######
487 # Global objects      #
488 ######
489 Gbl = Scenario()
490 Start(Gbl)

```

A.2 Sacrifice.py

```

1 ######
2 # EVOLIFE          http://evolife.telecom-paristech.fr  #
3 # Telecom ParisTech 2018           Jean-Louis Desselles #
4 # _____#
5 # License: Creative Commons BY-NC-SA                      #
6 ######
7
8 """
9     Study of the possibility of self-sacrifice being an ESS,
10    as first-order signal for patriotism.
11    Uses 'Exogenous.py' as Base (overloading useful functions)
12 """
13
14 from math import log
15 from random import randint, random, choice
16
17 import Exogenous as Base
18
19 import sys
20 sys.path.append('..')
21 sys.path.append('../..')
22
23 import Evolife.Ecology.Alliances as EA
24 from Evolife.Tools.Tools import percent
25

```

```

26
27 class Scenario(Base.Scenario):
28
29 ##### General initializations and visual display #####
30 ##### General initializations and visual display #####
31 #####
32
33     def genemap(self):
34         """ Defines the name of genes and their position on the
35         DNA.
36         """
37         return [('Patriot'), ('NonPatriot'), ('Demand'), (
38             'SelfSacrifice')]      # gene length (in bits) is read from
39             configuration
40
41     def update_positions(self, members, start_location):
42         """ Locates individuals on an 2D space
43         """
44         duplicate = members[:]
45         # sorting individuals by patriotism
46         duplicate.sort(key = lambda x: x.Patriotism)
47         for m in enumerate(duplicate):
48             m[1].location = (start_location + m[0], m[1].
49             SignalLevel)
50
51 ##### Life_game #### (1) Initializations #####
52
53     def prepare(self, indiv):
54         """ Defines what is to be done at the individual level
55         before interactions
56             occur - Used in 'start_game'
57         """
58         indiv.score(100, True) # Sets initial ("social") score
59         to 100
60
61         indiv.SignalLevel = 0 # Resets signal level - useful
62         for a non-genetic learning (future ?) version, not here
63         indiv.Reproduction_points = 0 ## Reset reproductive
64         points to 0
65
66         # friendship links (lessening with time) are updated (
67         # Erosion is null by default)
68         indiv.lessening_friendship((100 - self.Parameter('Erosion
69             '))/100.0)
70
71 ##### Life_game #### (2) Self-sacrifices #####
72 #####

```

```

65
66     def honoring(self, worshippers, nb_heroes):
67         """ Total social admiration (for heroes)
68             depends on the behaviors of all individuals in the 'honoring' game
69                 (second-order signal)
70         """
71         if nb_heroes == 0: return 0 # No heroes to honor
72         Offerings = 0
73         for Indiv in worshippers:
74             if Indiv.Patriotism == 0: # Indiv is not a patriot
75                 offering = Indiv.gene_relative_value('NonPatriot')
76             )
77                 if self.Parameter('DifferentialCosts') == 0: # MaxOffer version
78                     Indiv.score(-self.costHonor(offering))
79                     offering = min(self.Parameter('MaxOffer'), offering)
80                         # Non-patriot's offering is capped:
81                         # they already have other investments (in
82                         # the rival group, in themselves...)
83                         # If social demand is too high, this
84                         could push them
85                             # (perhaps similar to 'exclusive' signals
86                             : one can't wear clothes tied to two social classes at once,
87                             # or bear tatoos from two rival gangs...)
88                         else: # DiffCosts version
89                             Indiv.score(-self.costHonor(offering,
90 DifferentialCosts = True))
91                                 # Honoring is more costly for non-
92                                 patriots: investing time (for instance)
93                                     # in the group is less-interesting than
94                                     for patriots, since they aren't as committed
95                                         # to the group and thus stand to gain
96                                         less in relative terms purely in relation to the group
97                                         # (opportunity cost)
98                                         else:
99                                             offering = Indiv.gene_relative_value('Patriot')
100                                            Indiv.score(-self.costHonor(offering))

101
102     Indiv.SignalLevel += int(offering / self.Parameter('VisibleThreshold')) * self.Parameter('VisibleThreshold')
103         # Signal must surpass a threshold to be visible
104
105         # Visible signals become part of social admiration
106         # (In this linear version, number of heroes or
107         distance between members is not a factor)
108         Offerings += Indiv.SignalLevel
109
110     return Offerings

```

```

101
102
103     def costHonor(self, offering, DifferentialCosts = False,
104                  patriotism = 0):
105         basic_cost = offering * percent(self.Parameter(
106             'HonoringCost'))
107         if not DifferentialCosts:    # MaxOffer mode: non-patriots
108             face the same cost as patriots
109             return basic_cost
110         else:    # DiffCosts mode: NonPatriots face a premium for
111             honoring
112             return basic_cost + basic_cost * (1 - patriotism) *
113                 percent(self.Parameter('DishonestPremium'))
114
115 ###### Life_game #### (3) Social interactions #####
116 #####
117
118     def interact(self, receiver, Signalers):
119         """ Receiver selects the first Signaler it sees that
120             matches its demands
121             (and that accepts it as a follower)
122             """
123         Candidates = self.filter(Signalers, receiver.
124             gene_relative_value('Demand'))
125         if Candidates == []: return
126         for Signaler in Candidates:
127             if Signaler.followers.accepts(0) >=0:
128                 receiver.F_follow(0, Signaler, 0)
129                 break    # Signaler and receiver have accepted
130             each other as friends
131
132     def filter(self, Signalers, requirement):
133         """ Filters out signalers that don't meet the individual's
134             requirements in terms of visible patriotism """
135         AcceptableFriends = Signalers[:]
136         for Signaler in Signalers:
137             if Signaler.SignalLevel < requirement:
138                 AcceptableFriends.remove(Signaler)
139         return AcceptableFriends
140
141     def interact_old(self, indiv, Signalers):
142         """ Formation of friendship bonds
143             By honoring heroes (see 'honoring'), individuals
144             signal their patriotism
145             This signal is used by others to choose their friends
146             (keeping in mind this is crucial: see 'evaluation')
147             """
148         if Signalers == []: return

```

```

140         # The agent chooses the first Signaler it sees that
141         # matches its demands
142     ) # Signalers are not sorted by default
143         demand = indiv.gene_relative_value('Demand')
144         offer = indiv.SignalLevel
145         for Signaler in Signalers:
146             if Signaler == indiv: continue
147             if indiv.follows(Signaler): continue
148             if demand > Signaler.SignalLevel:
149                 continue # No available interesting Signalers
150             if offer < Signaler.gene_relative_value('Demand'):
151                 continue
152             if indiv.get_friend(1, Signaler, 1):
153                 return # Indiv and Signaler have become friends
154
155 ###### Life_game #### (4) Computing scores and life points #####
156 ###### Life_game #### (4) Computing scores and life points #####
157 ###### Life_game #### (4) Computing scores and life points #####
158
159     def evaluation(self, indiv):
160
161         indiv.score( + self.Parameter('JoiningBonus') * indiv.
162         nbFollowers() )
163
164         if indiv.Patriotism ==0 and random() < percent(self.
165         Parameter('NbTraitors')):
166             # indiv is a traitor who betrays its friends
167             for follower in indiv.followers:
168                 follower.score(- self.Parameter('DenunciationCost'
169 )) # The betrayed pay a cost
170                 indiv.score(+ self.Parameter('Judas'))
171             # Betrayer obtains payment for betrayal
172             else:
173                 for follower in indiv.followers:
174                     follower.score(+ self.Parameter('FriendshipValue',
175 )) )
176             return
177
178     def lives(self, members):
179         """ Converts alive members' scores into life points -
180         used in 'life_game'
181         """
182         if self.Parameter('EraseNetwork'): self.
183         reinitialize_network(members) # Reinitialize friendships (0
184         by default)
185         #AliveMembers = self.betrayal(members, Spare=self.

```

```

Parameter( 'DenunciationCost' ) # unused ( see under )
178     AliveMembers = members
179     if self.Parameter( 'SelectionPressure' ) == 0:
180         return
181     if not AliveMembers:
182         return
183     BestScore = max([ i.score() for i in AliveMembers ])
184     MinScore = min([ i.score() for i in AliveMembers ])
185     if BestScore == MinScore:
186         return
187     for indiv in members:
188         indiv.LifePoints = int( (indiv.score() - MinScore) *
self.Parameter( 'SelectionPressure' ) / (BestScore - MinScore)
)
189         # individuals obtain between 0 and SP points ,
depending on score relative to the population
190
191 def betrayal( self , members , Spare=True):
192     " Unused: for version where being betrayed leads to death
"
193     if Spare: return members
194     AliveMembers = members [:]
195     for i in members:
196         if i.Executed:
197             AliveMembers.remove( i )
198             i.LifePoints = -1
199     return AliveMembers
200
201 def reinitialize_network( self , members):
202     for i in members:
203         i.forgetAll()
204
205 #####
206 #####
207 #####
208 #####
209 # Adding patriotism
210
211 class Patriotic_Individual( Base.Individual , EA.Follower):
212     " Individuals now also have a patriotism phenotype "
213
214     def __init__( self , Scenario , ID , maxPatriotism = 100 , Newborn
=True):
215         self.Patriotism = randint(0 , 1)      # Binary patriotism
216
217         # Individual inherits from Exogenous
218         Base.Individual.__init__( self , Scenario , ID=ID , Newborn=
Newborn )
219

```

```

220         self.SignalLevel = 0
221         self.Executed = False    # Unused (see betrayal)
222
223     # Individual inherits from Follower module (which manages
224     # social relations)
224     EA.Follower.__init__(self, self.Scenario.Parameter('
224     MaxFriends'), Scenario.Parameter('MaxFriends'))
225
226
227 #####
228 #####
229 #####
230 # Overloading useful elements of Group module with local
231 # Patriotic_Individual class
232
232 class Group(Base.Group):
233     " In each group, patriotism ranges from 0 to group size (
233     simplification) "
234
235     def __init__(self, Scenario, ID=1, Size=100):
236         Base.Group.__init__(self, Scenario, ID, Size)
237
238     def createIndividual(self, ID=None, Newborn=True):
239         # calling local class 'Individual'
240         Indiv = Patriotic_Individual(self.Scenario, ID=self.
240         free_ID(), Newborn=Newborn)
241         # Individual creation may fail if there is no room left
242         self.Scenario.new_agent(Indiv, None) # let scenario know
242         that there is a newcomer (unused)
243         #if Indiv.location == None: return None
244         return Indiv
245
246     def reproduction(self):
247         """ Reproduction within the group (see Exogene) for
247         details
248             Calls local class 'Patriotic_Individual'
249             """
250             self.update_(flagRanking=True) # updates individual
250             ranks
251             for C in self.Scenario.couples(self.ranking):
252                 # Making of the child
253                 child = Patriotic_Individual(self.Scenario, ID=self.
253                 free_ID(), Newborn=True)
254                 if child:
255                     child.hybrid(C[0],C[1]) # Child's DNA results
256                     from parents' DNA crossover
257                     child.mutate()
257                     child.inherit_share(C[0],C[1], heredity = percent
257 (self.Scenario.Parameter('SacrificeHeredity')))
```

```

258         child.update() # Computes the value of genes, as
259         DNA is available only now
260         if self.Scenario.new_agent(child, C): # Let
261             scenario decide something about the newcomer (not used here)
262                 # Child is added to parents' children
263             lists
264                 C[0].addChild(child)
265                 C[1].addChild(child)
266
267
268
269 # Overloading useful elements of Population module with local
270 # Group class
271
272 class Patriotic_Population(Base.Population):
273     """ Defines the population of agents
274     """
275
276     def __init__(self, Scenario, Observer):
277         """ Creates a population of agents """
278         Base.Population.__init__(self, Scenario, Observer)
279         self.Scenario = Scenario
280
281     def createGroup(self, ID=0, Size=0):
282         """ Calling local class 'Group' """
283         return Group(self.Scenario, ID=ID, Size=Size)
284
285
286
287 if __name__ == "__main__":
288     print(__doc__)
289
290
291     # Global objects #
292
293     Gbl = Scenario()
294     Base.Start(Gbl, Patriotic_Population)

```

B | Mathematical demonstrations

B.1 Exogenous model

B.1.1 RemTH

B.1.2 $\beta(p)$

Disregarding the effects of natural selection (which are the same for all individuals here¹), an individual who bears a **SelfSacrifice** gene of relative value p , has a probability p of dying in his first year (before being able to foster any descendants), a probability $(1 - p) * p$ of dying at age 1 ... a probability $(1 - p)^n * p$ of dying at age $n < M$... and is certain to die at age M , should he or she reach it. His/her expected life span is thus:

$$ELS = p * 0 + (1 - p) * p * 1 + \dots + (1 - p)^{M-1} * (M - 1) + (1 - p)^M * M$$

Let $f: \mathbf{R} \rightarrow \mathbf{R}$ be the polynomial function defined by the expression:

$$f(x) = \sum_{n=0}^{M-1} p * (1 - p)^n * x^n + (1 - p)^M * x^M$$

By deriving f , one can note that:

$$ELS = f'(1) \tag{12}$$

$f(x)$ involves a geometric sum and can be simplified to (when $(1 - p) * x$ is different than 1):

¹And will therefore appear on both sides of an equation comparing the benefits of either strategy such as (2).

$$f(x) = p * \frac{1 - ((1-p) * x)^M}{1 - (1-p) * x} + ((1-p) * x)^M$$

For $x \neq \frac{1}{(1-p)}$ ($p = 1$ trivially yields $ELS = 0$):

$$f'(x) = p * \frac{-M(1-p)^M x^{M-1} * (1 - (1-p)x) + (1-p) * (1 - ((1-p)x)^M)}{(1 - (1-p)x)^2} + M(1-p)^M x^{M-1}$$

And thus:

$$\begin{aligned} f'(1) &= p * \frac{(-M(1-p)^M * p + (1-p) * (1 - (1-p)^M))}{p^2} + M(1-p)^M \\ f'(1) &= \frac{(1-p) * (1 - (1-p)^M)}{p} \end{aligned} \quad (13)$$

Combining these two expressions for $f'(1)$ proves equation (1). **Figure 15** shows $\beta(p)$ for p between 0 and 1. Factoring in natural selection, an individual's expected life span is therefore equal to: $\alpha * \beta(p) * M$.

B.1.3 $R_+(A)$

In **Selectivity** mode, individuals obtain reproductive potential R according to their **Reproductivepoints** RP , each individual obtains a rank k according to RP , and receives reproductive potential:

$$R = \frac{r}{2} * \left(\frac{S}{(S * k + N') * \log(1 + S)} + \frac{S}{(S * (k + 1) + N') * \log(1 + S)} * N' \right)$$

where $N' < N$ is the number of eligible parents (non-martyrs) and S is equal to *Selectivity* (the average reproductive potential over eligible parents being *ReproductionRate* r). When $N \gg 1$, R verifies:

$$R \in [R_{min}; R_{max}], \text{with } R_{max} \approx \frac{S * r}{\log(1 + S)} \text{ and } R_{min} \approx \frac{R_{max}}{1 + S} \quad (14)$$

With typical parameters ($S = 10$ and $r = 15\%$), we obtain: $R_{max} \approx 63\%$ and $R_{min} \approx 5,6\%$.

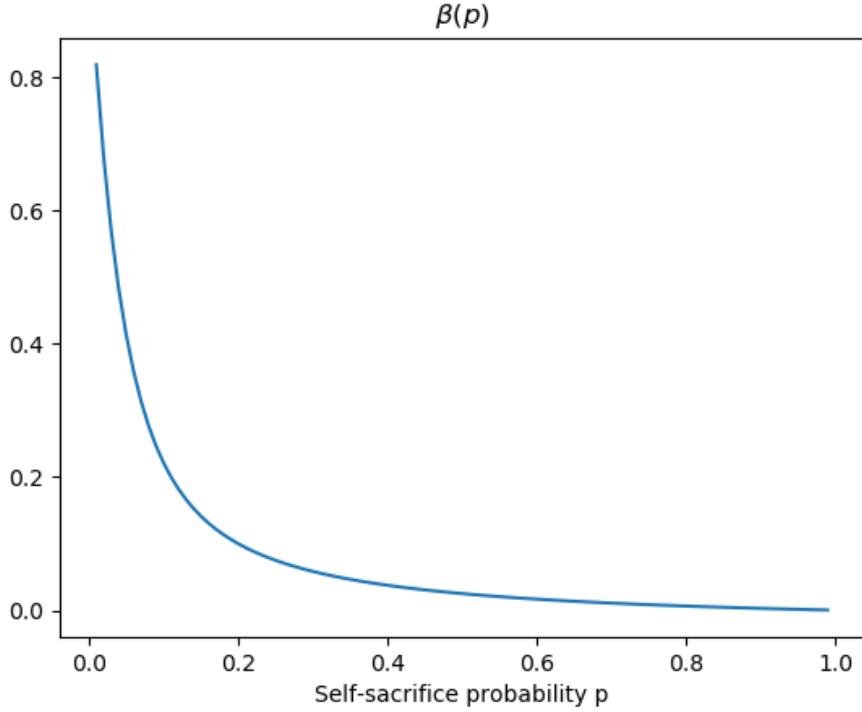


Figure 15: Loss of expected life-span due to self-sacrifice.

In a case where a negligible proportion f of individuals engage in self-sacrificial behavior (which is assured to end up in their martyrdom), their children each receive, on average:

- reproductive potential r , when $\frac{AN(1-P)}{fN*\alpha\beta(p)Mr} < RGT$;
- $R_+(A, f)$ otherwise, as seen in **Section 3.3.2**.

When $N \gg 1$, expected $R_+(A, f)$ is equal to the average between the "luckiest" ($k = 0$) and "unluckiest" child, which is approximately:

$$R_+(A, f) \approx \frac{S * r}{2 * \log(1 + S)} * \left(1 + \frac{N'}{S * fN * \alpha\beta(p)Mr + N'}\right)$$

$$R_+(A, f) \approx \frac{S * r}{2 * \log(1 + S)} * \left(1 + \frac{1}{\frac{S*fN*\alpha\beta(p)Mr}{(1-f)N*\alpha Mr+fN*\alpha\beta(p)Mr} + 1}\right)$$

$$R_+(A, f) \approx \frac{S * r}{2 * \log(1 + S)} * \left(1 + \frac{(1 - f) + f\beta(p)}{(S + 1) * f * \beta(p) + (1 - f)}\right)$$

Which yields, for $f \ll 1$ (neglecting terms of order 2 and above):

$$R_+(A, f) \approx \frac{S * r}{\log(1 + S)} * \left(1 - \frac{f * \beta(p)}{2}\right) = R_{max} * \left(1 - \frac{f * \beta(p)}{2}\right) \quad (4)$$

B.1.4 ESS

Using (2), we deduce that both envisioned strategies are equivalent if and only if (at this level of approximation):

$$\begin{aligned} & \frac{Sr}{\log(1 + S)} \left(1 - \frac{f\beta(p)}{2}\right) \beta(p) = r \\ \iff & f * \beta(p)^2 - 2 * \beta(p) + 2 * \frac{r}{R_{max}} = 0 \end{aligned} \quad (15)$$

This is a quadratic equation in $\beta(p)$, whose reduced discriminant δ verifies:

$$\delta = 1 - 2f \frac{r}{R_{max}} > 0 \text{ since } r < R_{max} \text{ and } f \ll 1$$

Since $\beta(p)$ is smaller than 1, the only possible solution is:

$$\beta(p) = \frac{1 - \sqrt{1 - \frac{2fr}{R_{max}}}}{f}$$

Yielding, for $f \ll 1$:

$$\beta(p) = \frac{r}{R_{max}} + \frac{1}{2} * \left(\frac{r}{R_{max}}\right)^2 * f \quad (5)$$

Since $\beta: [0, 1] \rightarrow [0, 1]$ is a bijection, as visible on **Figure 15**, this yields a unique $p(f) = \beta^{-1}(\frac{\log(1+S)}{S} + \frac{1}{2}(\frac{\log(1+S)}{S})^2 f)$ for a given f . With typical parameter values, and $f \leq 1\%$ we find $p \approx 9.3\%$, by zeroing in on p at a precision of .1%.

In addition, if we suppose that $(1-p)^{M+1} \ll (1-p)$, which is the case here, then we can calculate an approximation² for p :

²Which would have yielded $p \approx 9.4\%$.

$$p \approx \frac{1}{1 + M * (\frac{\log(1+S)}{S} + \frac{1}{2} * (\frac{\log(1+S)}{S})^2 * f)}$$

$$p \approx \frac{1}{1 + M * \frac{\log(1+S)}{S}} * (1 - \frac{\frac{1}{2} * (\frac{\log(1+S)}{S})^2}{1 + M * \frac{\log(1+S)}{S}}) * f \quad (6b)$$

In a potential self-sacrifice ESS involving a negligible proportion f of agents that lay down their life with probability p , then p can be thus approximated.