COGMASTER

SECOND-YEAR INTERNSHIP MEMOIR

Self-sacrifice as a social signal

Julien Lie

supervised by

Jean-Louis DESSALLES

June 7, 2019

Contents

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

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 Dessalles to study various evolutionary phenomena (https://evolife.telecom-paristech.fr/). Evolife has been used to study social signals in general (Dessalles, 2014) and to model language as a particular social signal (Dessalles, 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 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, 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 individuals 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 equlibrium 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

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

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 reffered 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 context [1] 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 reccurence 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 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 (Spartians 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

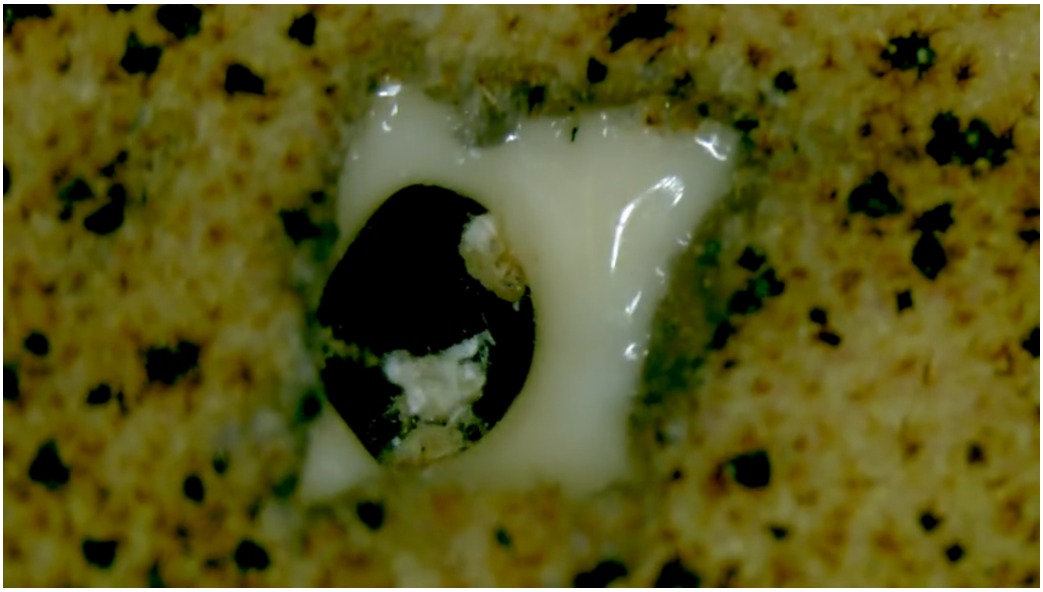


Figure 1: Soldier aphids using their body fluid to repair their gall. Kutsukake et al., via *lemonde.fr*.

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 gall [2], 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 mistake[3] – 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 adress question **(Q4)**.

[2]Contrary 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.

[3]Evolutionary 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 constrast 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 probabiltiy 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

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 phenomenons 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 assumptions[5] , 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

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

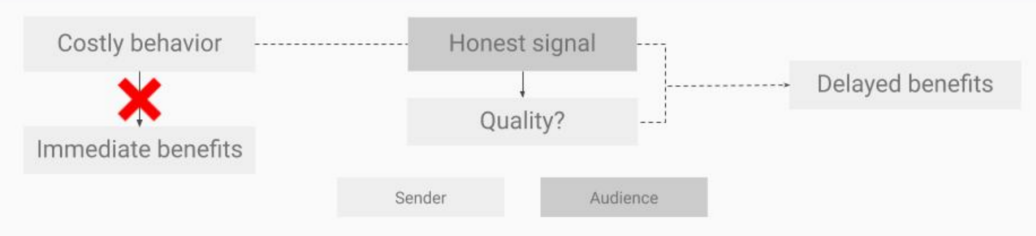
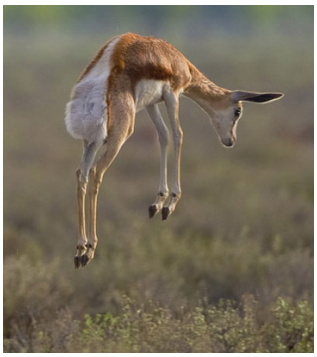


Figure 2: Signaling theory helps explain the emergence of costly behavior that is not followed by immediate benefits. Competition between senders helps garantee signal honesty, and thus their relevance to the audience.

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 costly [6]. 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,

[6] If 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 competition[7] (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.

[7]Which 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 so[8]. 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

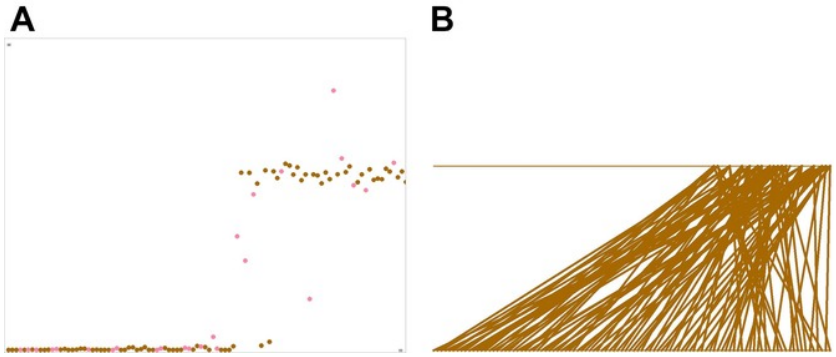


Figure 4: Emergence of an "elite" in the case of unconstrained competition (Dessalles, 2014; figures are his). Individual (dots) signal levels are represented on the left (**A**), and (asymmetrical) social bonds on the right (**B**).

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,

[8]Somewhat 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

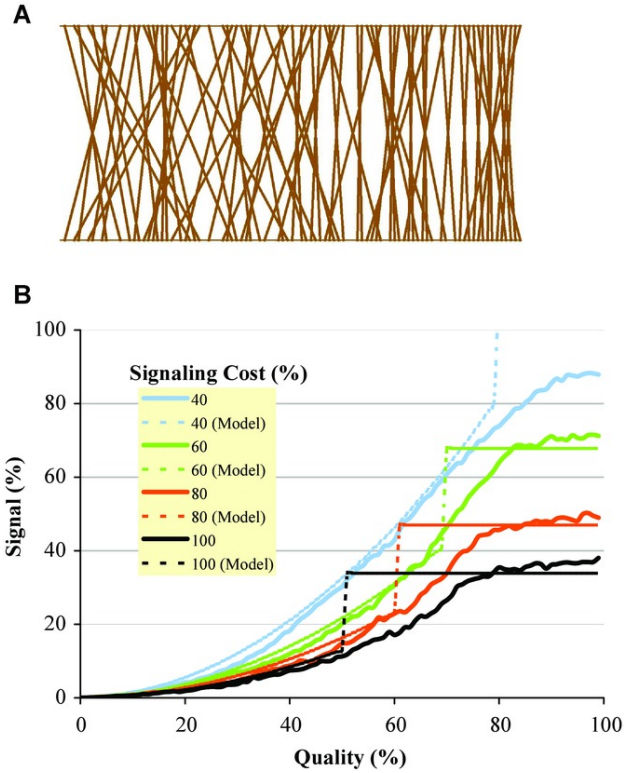


Figure 5: Emergence of generalized signaling (Dessalles, 2014; figures are his). Signaling value as function of individual quality is represented beneath (**B**), for differing costs; (symmetrical) social bonds above (**A**).

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 adressing question (**Q2**) within a signaling model:

Hypothesis 2 (H2): *Betrayal by friends is an important feature of intergroup 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 legitamely 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 to 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 adressing question (**Q4**), provided that martyrs are indeed revered and memorialized[9] (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,

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

[10]This 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 adress **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 Dessalles 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) sense[1]: 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**). 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 frequence of self-sacrifice, as per **(P1)**.

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

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

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).

This script serves as a basis for the full script, introduced below.

2.2.3 Arguments in favor of the proposed signals' honesty

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 that 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).

In extreme situations, competition can thus lead to large individuals 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, tentatively[3] suggesting that in such conditions, demand may allow extremely costly rights to emerge, which may warrant self-sacrifice – as mediated by the proximate feeling of "extreme" identity fusion.

2.2.4 Actual script structure

This script (*bfSacrifice.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 individuals 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

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

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. As argued above (see **Section 2.2.3**)), this could be the case if non-patriots honoring potential is capped ('MaxOffer' parameter) or if they face a signaling premium ('DishonestPremium').

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 ther 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 equlibrium 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 simluations

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 overwise), 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

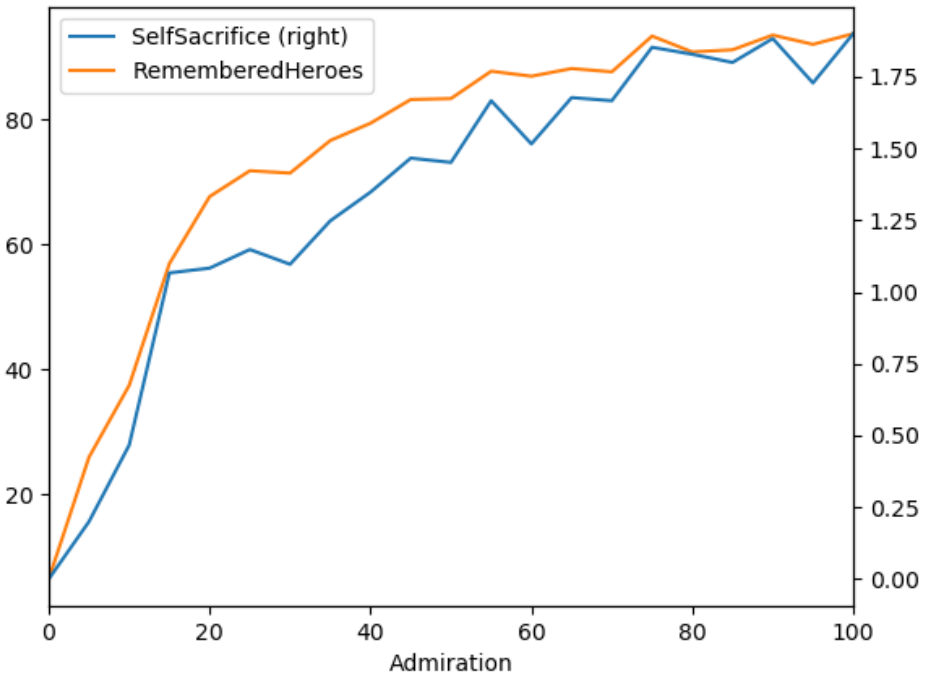
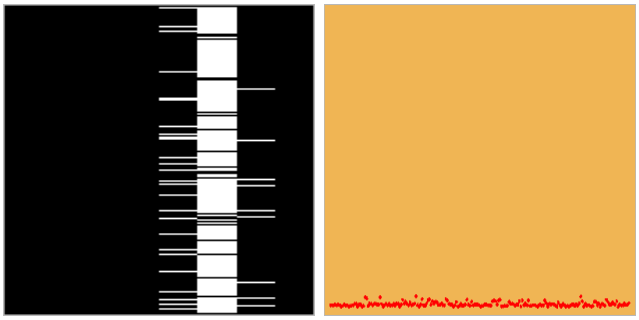


Figure 6: **SelfSacrifice** and **RememberedHeroes**, as a function of *Admiration* (typical parameter values).

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**.



(a) Genetic values (b) Reproductive points

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 approximately[1], it can best be approached by a function of this form with parameters $C = 95$ and $\tau = 20$, as visible on **Figure 8**.

[1]By 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.

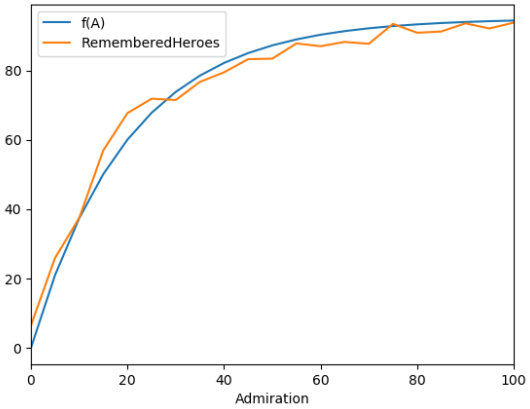
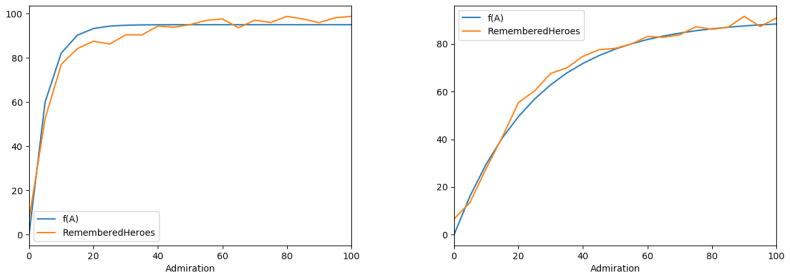


Figure 8: **RememberedHeroes** and corresponding optimal f as functions of *Admiration*, for $RGT = 10$ ($C = 95, \tau = 20$).



(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.

[2]This 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

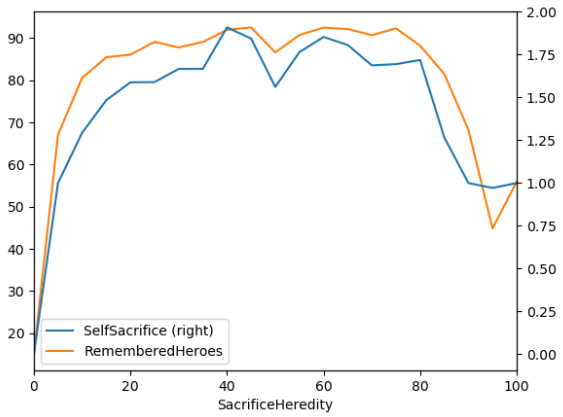


Figure 10

3.2.4 Variation with *ReproductionRate* r

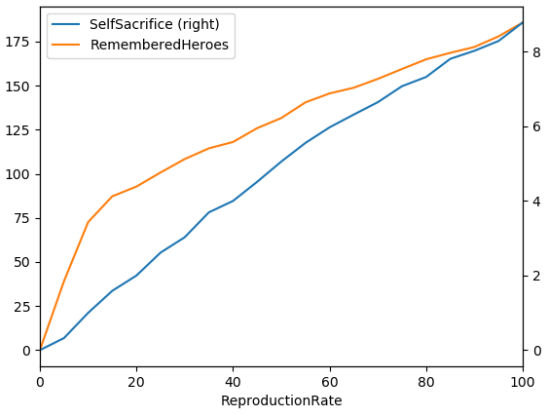


Figure 11

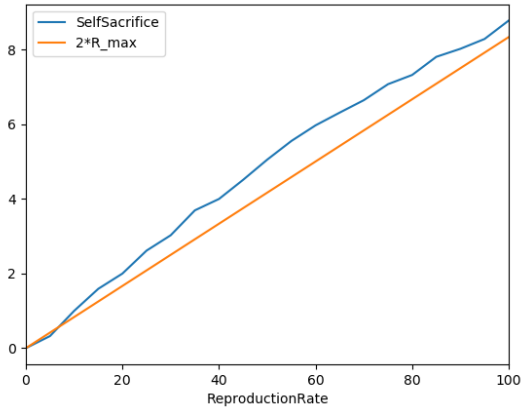


Figure 12

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 **Appedix B.1.2**):

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

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 large*[3] 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-order [4] (neglecting the effects on subsequent generations) characterization of equivalence between both strategies can be written as:

$$R_+(A, f) * \beta(p) = r \tag{2}$$

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:

[3] We 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.

[4] Note 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:

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

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*):

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

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

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

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

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

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

3.3.4 Brief Discussion

p = + function of S = duh // - of M : more to sacrifice But **no r**, no A $---> shouldbecapturedbyf....$: $OKforAinf_max...$ 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«1, all receive the same...

+ h: not here, vu le modele

+ lim for JLD : does not depend on nb of heroes... -> visibility log(nbheroes... ?)

3.4 Two-tier model simulation outputs

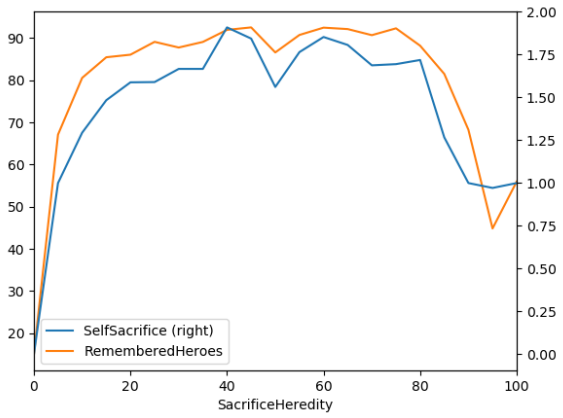


Figure 13

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 emerge[5], 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):

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

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:

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

3.5.2 Honoring and social score

This is motivation to investigage the possibilty of an honoring equilibrium. Honoring h and demand d serve as bases for establishing friendships, which are mutually beneficial (as controlled by *FriendshipValue* F). However, intergroup 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 characterzied 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:

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

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

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

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):

$\mathbf{E}(\Delta_P(d,h))$	$d \leq s_{\overline{P}}$	$d > s_{\overline{P}}$
$d \leq s_P$	$\frac{k}{2} * (2F + P - t * DC)$	$k * (F + P)$
$d > s_P$	$k * (F - t * DC)$	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 $\mathbf{E}(\Delta_{\overline{P}}(d,h)) \geq \mathbf{E}(\Delta_P(d,h))$[6]. For both, the 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:

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

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 $s_{\overline{P}}$ 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 $d > MO$ then $d \leq MO$. 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 $s_{\overline{P}} \geq s_P$[8]: the optimal situation represented by **Table 2**'s top right cell is unattainable, and, on average, patriots stand to lose from friendship.

[6] At least when h is small - see after

[8] And $d_{\overline{P}} \leq d_P$

Conversely, if $d > MO$, patriots either obtain no friends or only patriot friends and thus stand to benefit. Any stratetgy 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 stratetgy is thus MO^+, following the same reasoning as above.

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

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

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

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:

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

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

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

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

3.6 Second sacrifice model

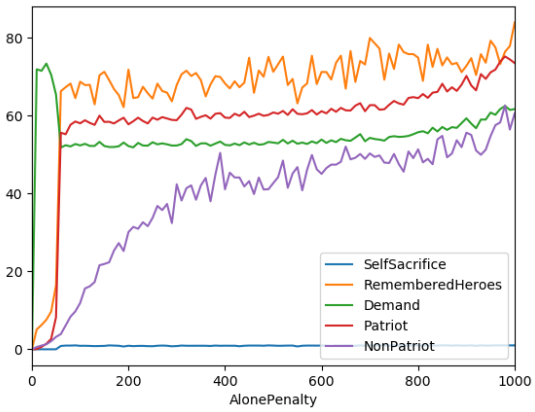


Figure 14

3.7 Limitations

4 | Perspectives

A | Python stuff

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 here[1]), 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:

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

Let $f\colon \mathbf{R} \to \mathbf{R}$ 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):

[1] And 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:

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

$$f'(1) = \frac{(1-p) * (1 - (1-p)^M)}{p} \tag{13}$$

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:

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

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 $N >> 1$, R verifies:

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

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

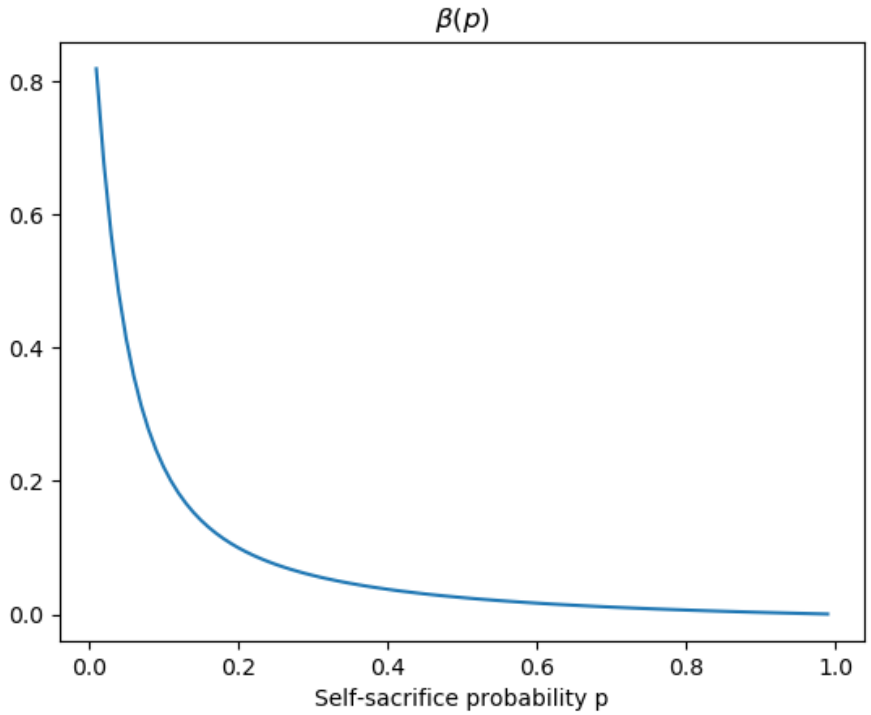


Figure 15: Loss of expected life-span due to self-sacrifice.

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

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

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

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

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

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

B.1.4 ESS

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

$$\frac{Sr}{log(1+S)}(1 - \frac{f\beta(p)}{2})\beta(p) = r$$

$$\iff f*\beta(p)^2 - 2*\beta(p) + 2*\frac{r}{R_{max}} = 0 \quad (15)$$

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

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

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 $f << 1$:

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

Since $\beta\colon [0,1] \to [0,1]$ 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 $(1-p)^{M+1} << (1-p)$, which is the case here, then we can calculate an approximation[2] for p:

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

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

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

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.