

# Properties of the epistemic-evolutionary cost function

## Some lessons from the intelligence of life

$$p = \mathfrak{b}$$

Bayesian methods labs

### Abstract

For a long time it was believed that the evolution of cooperation was conditioned by a dilemma. However, the evolutionary selection processes of life forms, by sequence of survival and reproduction rates, are like the selection processes of hypotheses in probability theory, by sequences of predictions, multiplicative in nature. Because in them the impacts of losses are stronger than those of gains (a zero in the sequence produces an irreversible extinction), the variants that flourish are hypotheses or life forms that reduce fluctuations by individual diversification (*epistemic property*), by cooperation (*evolutionary property*), by cooperative specialization (*speciation property*), and by cooperative heterogeneity (*ecological property*). There is no such dilemma because in multiplicative selection processes, defecting individuals increase the fluctuations of the cooperators on whom they depend, increasing therefore their own fluctuations and negatively affecting their own long-term growth rate (*mayor property*). While certain circumstances may favor defection in the short term (*minor property*), the emergence of higher level cooperative units is a permanent phenomenon in the history of life (major evolutionary transitions). The evidence is our own life, which depends on at least four of these levels for survival: the cell with mitochondria, the multicellular organism, the social system and the ecological community. Moreover, in probability something similar occurs: sets of individual hypotheses form variables, relationships between variables form causal models, and several models form theories. In the history of human beings, the cultural transition had obvious positive effects: before epistemic cooperation (transmission of knowledge) we were in serious danger of extinction; then we were able to occupy all the ecological niches of the earth as no other terrestrial vertebrate. In recent history, certain junctures produced the emergence of colonialism, an era of genocide and massive loss of cultural diversity. Despite all the advances, metropolitan science has not been able to compensate for the loss of millenary knowledge, and the current ecological crisis continues to deepen. In the long term, only variants capable of reducing fluctuations through individual diversification and through cooperative heterogeneous specialization survive.

Science is a human institution that has a claim of truth: reach intersubjective agreements with intercultural (or universal) validity. The formal sciences (mathematics, logic) reach these agreements by deriving theorems within closed axiomatic systems. However, the empirical sciences (from physics to the social sciences) must validate their propositions within open systems that always contain some degree of uncertainty. Is it then possible to reach intersubjective agreements (“truths”) in the empirical sciences if it is inevitable to say “I don’t know”? Yes. In short, we can avoid lying: not say more than what is known while incorporating all the available information.

For example, suppose we know that there is one gift hidden behind a box among three. Where is the gift? If we choose any of the boxes we would be stating more than we know, because we have no information that would make us prefer any of them. In this case, we will agree on the need to divide



Figure 1. Two belief distributions. Maximizing uncertainty allows reaching a first intersubjective agreement.

belief into equal parts. By maximizing uncertainty we avoid saying more than we know, allowing us to reach a first intersubjective agreement in contexts of uncertainty! Great. But then, how do we preserve this intersubjective agreement if we receive new information?

The logic of uncertainty (probability theory) has been repeatedly derived from different principles (axioms), always arriving at the same two simple rules. The *sum rule* ensures that we do not lose belief when we distribute it among mutually contradictory hypotheses: by adding up how much we assign to each hypothesis, we recover the initial 100%. And the (conditional or) *product rule*, guarantees the coherence of beliefs with the available information: we preserve the previous belief that is still compatible with the new data (and the surviving belief is our new 100%).

$$\begin{array}{c}
 \text{Sum rule:} \\
 P(A) + P(\text{not } A) = 1 \quad , \quad \underbrace{P(\text{Hypothesis}|\text{Data})}_{\substack{\text{Conditional belief} \\ \text{New belief}}} = \underbrace{P(\text{Hypothesis}, \text{Data})}_{\substack{\text{Joint prior belief} \\ \text{compatible with the data}}} / \underbrace{P(\text{Data})}_{\substack{\text{Prediction of the data} \\ \text{total compatible belief}}}
 \end{array}$$

These are the rules of probability for updating beliefs in contexts of uncertainty, there are no other rules. For example, suppose we are now allowed to choose one of the boxes and then someone gives us a hint, opening a box other than the one we chose in which the gift is not in (Figure 2). With this prior information we can define a causal model of the hint (Figure 2a). Before updating the belief about the position of the gift with information from the hint, we need the initial belief that arises from maximizing uncertainty by dividing the belief equally by the alternative paths of the causal model (Figure 2b). And finally, the new belief arises from preserving the initial belief that remains compatible with the data (Figure 2c).

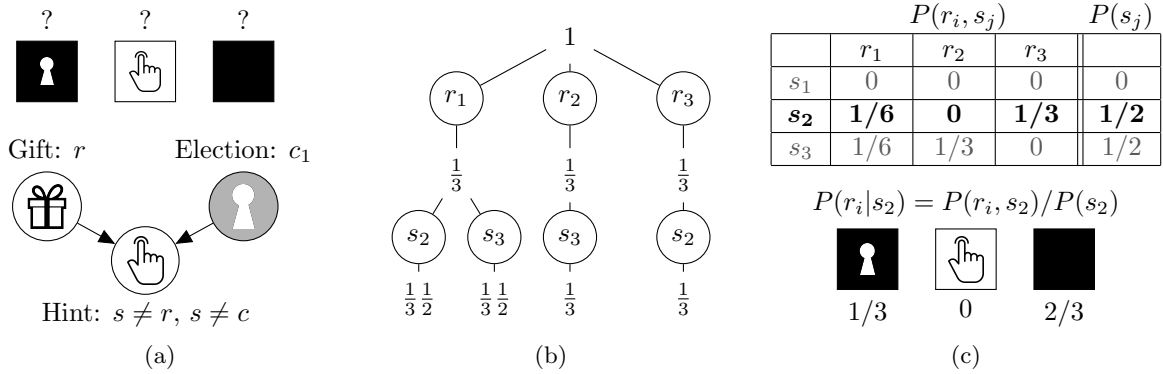


Figure 2. Intersubjective agreement given all the available information (causal model and data). (a) Causal model in which the hint  $s$  can neither point to the chosen box  $c$  (with the lock) nor to the box where the gift is located  $r$  (the hidden hypothesis):  $s \neq c$  and  $s \neq r$ . (b) Before knowing the hint, we define the prior belief by maximizing the uncertainty given the causal model (dividing the belief equally at each bifurcation produced by the parallel universes). (c) The new belief is nothing more than the initial belief (table) that remains compatible with the data  $s_2$ , expressed as 100%.

Unlike ad-hoc approaches that select a single hypothesis (e.g. by maximum likelihood), the strict application of probability rules (Bayesian approach), by believing at the same time in mutually contradictory hypotheses (A and not A), allows surprise, the only source of information, to be the filter of prior beliefs. In general, if we have  $\text{Data} = \{\text{data}_1, \text{data}_2, \dots\}$ ,

$$\underbrace{P(\text{Hypothesis}, \text{Data})}_{\substack{\text{Initial belief compatible} \\ \text{with the data}}} = \underbrace{P(\text{Hypothesis})}_{\substack{\text{Initial intersubjective} \\ \text{agreement}}} \underbrace{P(\text{data}_1|\text{Hypothesis})}_{\text{Prediction 1}} \underbrace{P(\text{data}_2|\text{data}_1, \text{Hypothesis})}_{\text{Prediction 2}} \dots$$

If the prediction of the observed data is 1 (null surprise), then we preserve all our prior belief in that hypothesis. If the prediction of the observed data is 0 (total surprise), then the hypothesis becomes false forever.

The value of truth is not abstract, it is pragmatic. The “lie” is a cross-cultural concept and its negation “not to lie” is a principle present in all societies of the world not because there is a striking coincidence between different particular moral criteria, but because truth has a concrete practical value. The selection processes of hypotheses in probability theory, by sequences of predictions, are like the evolutionary selection processes of life forms, by sequence of survival and reproduction rates, multiplicative in nature.

Because in them the impacts of losses are stronger than those of gains (for example, a single zero in the sequence produces an irreversible extinction) there is an advantage in favor of variants (hypotheses or life forms) that reduce fluctuations.

Since its origin, life has acquired an extraordinary complexity in terms of diversification, cooperation, specialization and heterogeneity. To see why, let's see what happens with a multiplicative process. For example, suppose a bookmaker offers payouts  $Q_c > 0$  and  $Q_s > 0$  when a coin comes out Heads and Tails respectively, and that individuals are forced to gamble at each time step all their resources, assigning a proportion  $b_c = b$  to Heads and  $b_s = 1 - b$  to Tails (gambling only a part of the resources can be shown to have an equivalent solution). If our initial resources are  $\omega_0$ , the resources we obtain with a bet  $b \in [0, 1]$ , after one Heads and one Tails, is updated as a product of our results (see equation 1).

$$\omega_2(b) = \underbrace{\omega_0 \overbrace{b Q_c}^{\text{Cara}}}_{\omega_1(b)} \overbrace{(1-b) Q_s}^{\text{Seca}} \quad \frac{\omega_T(b)}{\omega_T(d)} = \frac{\omega_0 (b Q_c)^{n_c} ((1-b) Q_s)^{n_s}}{\omega_0 (d Q_c)^{n_c} ((1-d) Q_s)^{n_s}} \quad (1)$$

And if we compare the resources of two different bets  $b \neq d \in [0, 1]$  after obtaining  $n_c$  Heads and  $n_s$  Tails, at time  $T = n_c + n_s$ , we find that the relative value of one bet with respect to the other is independent of the payouts  $Q_c$  and  $Q_s$  offered by the bookmaker! Yes, we are saying that we can decide the bet without knowing the payouts. In general, we want the bet  $b$  that maximizes the growth rate  $r(b)$ ,

$$\omega_0 r(b)^T = \omega_0 (b Q_c)^{n_c} ((1-b) Q_s)^{n_s} \quad \underbrace{\arg \max_b}_{\text{Optimal bet}} r(b) = \underbrace{n_c/T}_{\text{Observed frequency}} \quad (2)$$

Indeed, no matter the payouts offered by the bookmaker, the optimal bet  $b^*$  is the one that divides the resources in the same proportion as the observed frequency  $f = n_c/T$ , which in the long run tends to the typical frequency  $\lim_{T \rightarrow \infty} n_c/T = p$ . This is the *epistemic property* that allows probability theory to acquire knowledge about the world, because evaluating individual hypotheses on the basis of the product of surprises produces an advantage in favor of hypotheses that diversify resources (probabilities) in the same proportion as the observed frequency.

Similarly, the evolutionary selection process by sequences of growth and reproduction rates produces the emergence of life forms that reduce fluctuations through individual diversification. For example, suppose the bookmaker offers  $Q_c = 3$  for Heads and  $Q_s = 1.2$  for Tails. If we know that the coin is unbiased the best we can do in the long run is to divide the resources equally,  $b = p = 0.5$ . In this case, however, we cannot win this game individually because Growing 50% with Heads ( $1.5 = Q_c b$ ) is not enough to compensate for the 40% decline we suffered with Tails ( $0.6 = Q_s (1-b)$ ). Even with the optimal behavior, individually the resource trajectories fall at a rate close to 5%, since  $r(b) \approx 0.95$  (Figure 5).

Although individual diversification is not enough to win this game, life found strategies to flourish in cases like these, further reducing individual fluctuations through cooperation. This is the *evolutionary*

	$\omega_0$	$\Delta$	$\omega_1(b)$	$\Delta$	$\omega_2(b)$
A no-coop	1	1.5	1.5	0.6	<b>0.9</b>
B no-coop	1	0.6	0.6	1.5	<b>0.9</b>
A coop	1	1.5	1.05	0.6	<b>1.1</b>
B coop	1	0.6	1.05	1.5	<b>1.1</b>

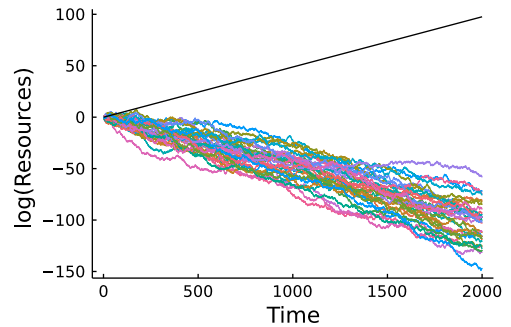


Figure 5. Cooperation allows individuals to reach growth rates they would never achieve alone. In the table (left)  $\Delta$  represents the change in resources experienced by individuals at random, with  $b = p = 0.5$ . In the figure (right), the colored curves are the trajectories of the resources playing individually, and the black line is the trajectory of the individual resources playing in big cooperative groups.

*property*. In our example, if at each time step individuals redistribute resources equally before betting, the growth rate of all individuals becomes positive.

While playing individually we were forced to bet ignoring the payouts offered by the bookmaker, once cooperation emerges it is no longer necessary to reduce fluctuations by individual diversification and an advantage in favor of cooperative specialization appears, making it possible to benefit from the best available option, further increasing the growth rate of all individuals in the cooperative group. This is the *speciation property*. In a cooperative group of size  $N$ , where resources are divided equally, the growth rate of individuals becomes the average of the resources after  $m_c$  and  $m_s$  individuals obtain Heads and Tails respectively, with  $m_c + m_s = N$  (see equation in Figure 6). As the group gets larger, the optimal

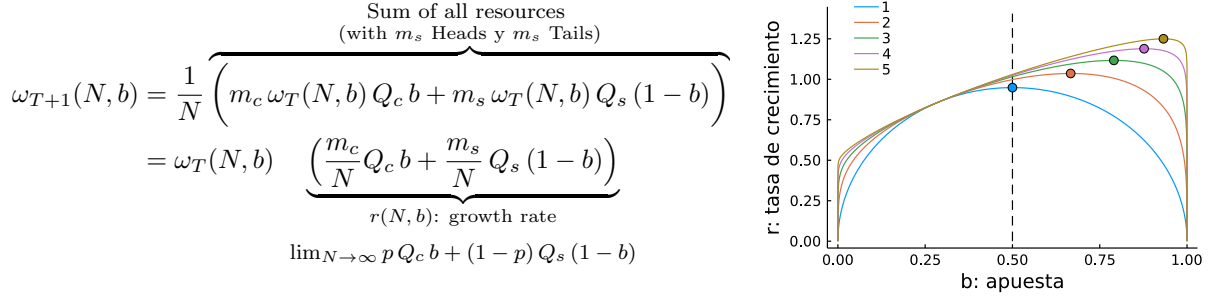


Figure 6. In the equation we see the growth rate in cooperative groups. In the figure we see the growth rate for groups of size 1 to 5 (colors), for all possible bets (x-axis). As the group size increases, the optimal bet becomes more specialized towards the option best paid by the bookmaker. When the size of the group tends to infinity, the optimal setting is to bet everything on the best option.

bet becomes “specialized” to the best available option (Figure 6).

In this case, all individuals are subject to stochastic processes that, although independent, have for all of them the same probability that does not change over time. In nature, on the contrary, we are exposed to spatio-temporal variabilities, such as the seasons. To give an example, suppose the coin oscillates between two probabilities of coming up Heads,  $p_1 = 0.2$  and  $p_2 = 0.6$ . At odd time steps, the coin has probability  $p_1$  in hemisphere A and  $p_2$  in hemisphere B. And vice versa in the even time steps. In cases like these, while specialization remains valuable for cooperative groups, the variability of the process produces an additional advantage in favor of groups with greater internal heterogeneity. This is the *ecological property*. If  $h$  is the proportion of individuals that are in hemisphere A, the growth rate of individuals in a large specialized cooperative group ( $\lim_{N \rightarrow \infty} r(h)$ ) is maximized when the distribution between hemispheres is in equal parts, with  $h = 0.5$  (figure 7).

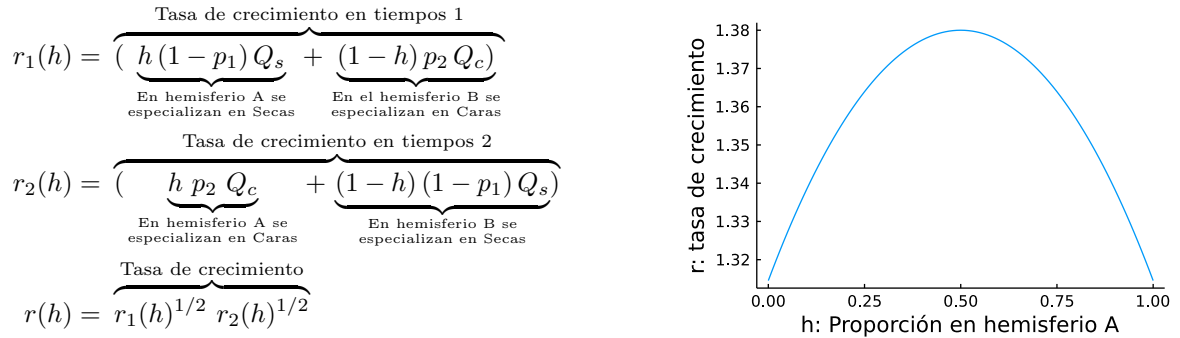


Figure 7. In the equation we see the growth rate as a function of the proportion in hemisphere A,  $h$ , of a large group specialized in the best option, in which individuals apply the optimal bet corresponding to the stage of the cycle they experience. In the figure we see that the growth rate of the individuals in the group is maximized when the heterogeneity among individuals is maximized.

These properties of multiplicative selection processes have produced throughout the history of life a

series of evolutionary transitions in which entities capable of self-replication, after the transition, became part of indissoluble cooperative units. Our own life itself depends on various levels of cooperation with heterogeneous specialization, without which we are not able to survive: the union of our cells with mitochondria and the emergence of organelles; our multicellular organism and the emergence of organs; our society and the emergence of roles and groups; the coexistence between species and the emergence of ecosystems.

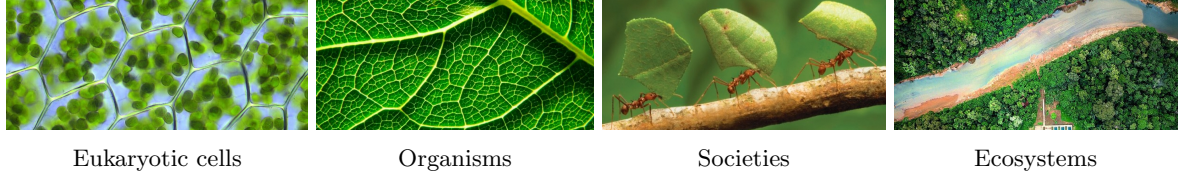


Figure 8. Major evolutionary transitions. The emergence of higher level units in life.

For a long time it was believed that for cooperation to evolve, the so-called “prisoner’s dilemma” had to be solved first. In it, cooperating involves a cost  $c$  for the other individual to receive a benefit of value  $v$ , with  $v > c$ , and defecting means refusing to cooperate and carries no cost, (see matrix in Figure 10). In our betting example, there would be a temptation to defect: to ceasing to contribute to the common fund while continuing to receive the common fund share. If evolution were subject to a prisoner’s dilemma, then defecting in entirely cooperative groups should offer to the individual a higher growth rate than cooperating. However, under multiplicative selection processes it is not possible to take advantage of defection because, by increasing the fluctuations of cooperators on which it depends, it also increases its own fluctuations, negatively affecting its own long-term growth rate. In Figure 10 we show the growth rate of individuals in cooperative populations with 0, 1 and 2 defectors.

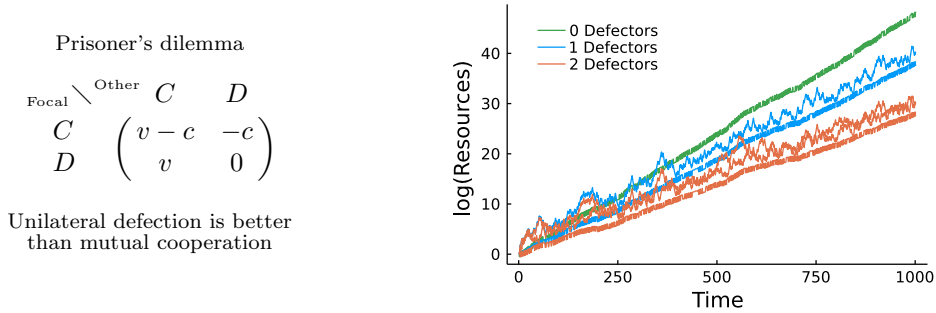


Figure 9. In the equation we see the payoff matrix of the prisoner’s dilemma. In the figure we see the growth rate of individuals within cooperative groups (colors) with 0, 1 and 2 defectors. The curves with the highest variability are the defecting individuals, the rest are the cooperators. The defecting individuals have a lower growth rate than they had when they were cooperating (*mayor property*), even though they might initially make a profit (*minor property*).

Similarly, higher-level cooperative units also emerge in probability theory: individual hypotheses form variables, variables form causal models, and sets of models form theories. The higher level hypotheses reduce fluctuations by making predictions with the contribution of all the lower level hypotheses that compose it. For example, the prediction of a model, by the sum rule, is the weighted average of the predictions of all the hypotheses that compose it.

$$P(d_{n+1}|d_1 \dots d_n, \text{Model}) = \sum_h^{\text{Hypothesis}} P(d_{n+1}|d_1 \dots d_n, h, \text{Model})P(h|d_1 \dots d_n, \text{Model})$$

Thus, by making the prediction with the contribution of all the lower level hypotheses, the models reduce the chances of introducing extreme values (a zero) in the multiplicative selection processes to which they are subject, producing better results than would be obtained with any of them individually.

Moreover, in the history of human beings, epistemic cooperation (transmission of knowledge between individuals) allowed the emergence of worldviews, such as science and culture, where highly specialized

paradigms in particular domains coexist. The positive effects are clear. Before the cultural transition, we were in serious danger of extinction, as evidenced by the low diversity of the human genome, even compared to the closest hominins. But when knowledge, that previously had to be rediscovered individually, became a common resource passed on from generation to generation, we began to be able to occupy all the ecological niches of the earth as no other terrestrial vertebrate had done before. Organized in small nomadic societies, we managed to walk from Africa to South America (arrows in Figure 10a).

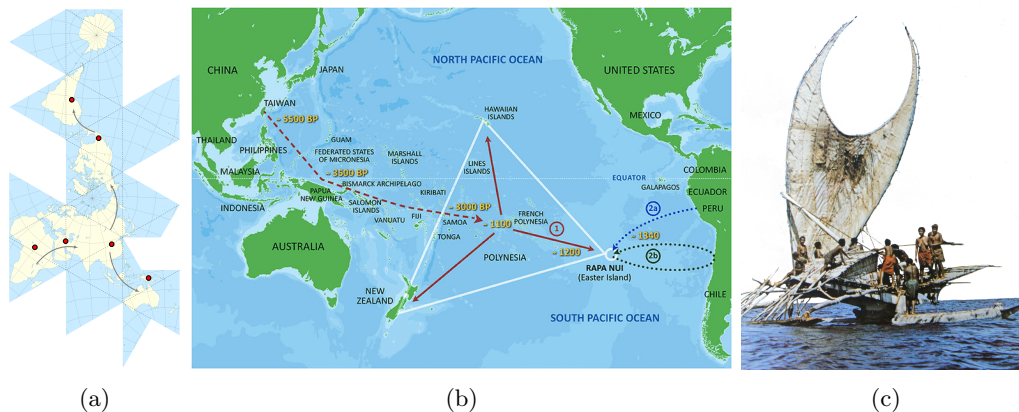


Figure 10.

In turn, the ecological reciprocity through prolonged symbiotic interaction with plants and animals led to the independent emergence of agriculture in sub-Saharan Africa, Middle East, China, Oceania, North America and South America (red points in Figure 10a). With it came an increase in population that intensified the processes of cumulative cultural evolution, making these regions the main technological centers of humankind.

During the 1400s, prosperous societies flourished all over the world. The Pacific Ocean was fully occupied, and there had already been exchanges between Oceania and South America, which is reflected in the genetics of the current populations (figura 10b). In the garden of human genetic and cultural diversity, sub-Saharan Africa, the Bantu society developed, among others. But China was the main productive and technological center of the world after 2 millennia. By 1400, the Arab world traded its products from the Atlantic Ocean in Spain to the Pacific Ocean in the Philippines. After conducting the explorations led by Zheng He (1405-1433) around the world, China recognizes that there is nothing interesting beyond its borders.

At the same time, the massive destruction of cultural diversity previously produced by the Roman Empire in its environment had led Western Europe, against the rest of the world, into a long process of cultural involution and internal violence known as the “Middle Ages”. Breaking the cooperative pact had concrete negative consequences for this society. At this stage, the criterion of authority as the foundation of “authentic knowledge” became commonplace within feudal society. A unbridled penal system was reborn from the so-called *libris terribilis* and the institutions inherited from the Western Roman Empire began to regulate community relations of sexual reproduction in a more detailed way than with private property. The war against women is definitively formalized with the publication of the *Malleus maleficarum* in 1484, which would be the second best seller after the Bible for the next 200 years.

Even there is an evolutionary disadvantage of defection, certain conjunctures can produce its emergence in the short term (*minor property*). Thus, the coincidence of a set of events, triggered by the massive feudal migration to America, suddenly put this historically marginal society in a situation of global privilege. The diseases transmitted by the feudal explorers eliminated, at least, 2/3 of the American population. However, there was no geopolitical shift until explorers discovered the silver mountain of Potosi in 1546, a metal that China had recently incorporated as one of its official currencies. Thanks to it, a quarter of a century later Western Europe finally managed to break its isolation (battle of Lepanto 1571) and begin the long cycle of importing foreign technology, mainly from China.

Until the early 1800s Western Europe continued to face a trade deficit with China, which was financed since the mid-1700s by the opium drug trade despite China’s prohibition. The consequences were more



severe when a cheaper and more potent opium mixture was developed in 1818. The number of addicts became alarming, and in 1839 China made the mistake of declaring war against the British narco-state in its own territory. The results were terrible: China lost 1/5 of its population and was mired for a

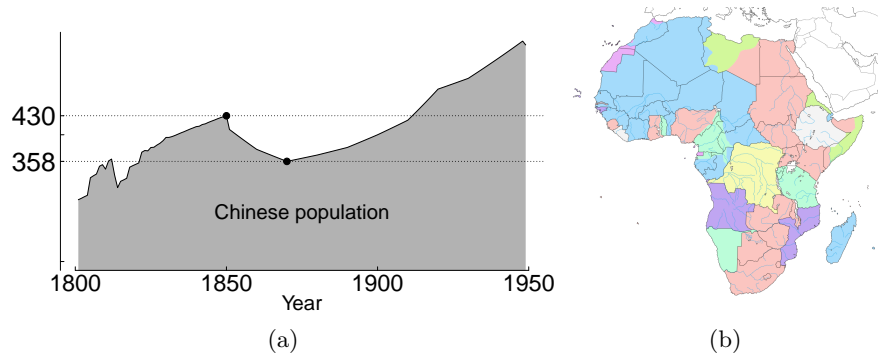


Figure 11. (a): China's defeat against the British narco-state on its own territory had terrible consequences. (b): It was not until 1850 that the colonization of continental Africa and the vast territories of the still autonomous Americas began.

century by foreign invasions.

When Western Europe finally establishes itself as the global geopolitical center, after the defeat of China, the era of genocides and massive loss of global cultural diversity begins: the violent occupation of continental Africa and the still autonomous territories of the Americas, among the main ones. The criterion of authority globalized during colonial-modernity as a criterion of universality limited to white men opened the door to cultural and ecological arbitrariness. In a world of weakened communities, the individualistic and instrumental worldview of our time expands: life and people as commodities. Despite all the advances, metropolitan science was not able to compensate for the loss of millenarian knowledge caused by colonial-modernity and as a consequence of the massive loss of global cultural diversity, we are currently experiencing a serious and deepening ecological crisis.

The advantage in favor of diversification and cooperation is not only theoretical, its breakdown has consequences for life and knowledge. Just as selecting a single hypothesis has known negative consequences in probability (overfitting), the imposition of a single type of society is having increasingly obvious ecological consequences for life. Ultimately, in the long term, only the variants capable of reducing fluctuations through diversification and cooperation survive.

Thus, the experience accumulated by the most diverse communities in the world has led, independently, to a universal obligation to give and receive, and to the development of technologies of reciprocity that maintain cohesion by re-activating community bonds through (festive or coercive) exchange rites. The right to cultural diversity is based on the practical fact that ecologically adapted cultural knowledge evolves only through the experience accumulated by autonomous peoples in the course of their own history. Just as the Bayesian approach adapts knowledge by believing at the same time in A and non-A, a Plurinational society adapts to life through reciprocal coexistence between autonomous communities.

 Bayes  
Plurinacional