The Most Demanding Moral Capacity: Could Evolution Provide Any Base?

La capacidad moral más exigente: ¿Podría la evolución proporcionar alguna base?

Teresa Bejarano[1]
University of Sevilla
Sevilla, Spain
tebefer@us.es
ORCID 0000-0002-0037-549X

Abstract: The attempts to make moral and evolution compatible have assimilated moral capacity either with complex self-control in favour of one's own goals or with spontaneous altruism. Those attempts face an easy problem, since those two senses of moral are adaptively advantageous resources. But let us focus on the decisions made in favour of another person which the subject, when making them, feels are contrary to his own goals: Could a base for this capacity arise in evolution, however poor and weak? I propose that such base, while it is not an adaptive advantage but quite the opposite, arises from the convergence between two abilities which in their respective origins were adaptively very advantageous: the advanced mode of 'theory-of-mind' (ToM) and inner speech.

Keywords: others' mental contents; speech directed to oneself; spontaneous altruism; Advanced Theory of Mind; vicarious expectations.

Resumen: Los intentos de hacer compatibles la moral y la evolución han asimilado la capacidad moral con el autocontrol complejo en favor de las metas propias o con el altruismo espontáneo. Esos intentos se enfrentan a un problema fácil, puesto que esos dos sentidos de la moral son adaptativamente ventajosos. En cambio, las decisiones que van contra las propias metas de uno son desventajosas. A pesar de ello, ¿pudo surgir en la evolución una base, por pobre y débil que fuera, para esta capacidad? Propongo que tal base, si bien no es una ventaja adaptativa en principio, sino más bien lo contrario, surge de la convergencia entre dos habilidades que en sus respectivos orígenes sí eran muy ventajosas adaptativamente: el modo avanzado de la teoría de la mente (ToM) y el habla interior.

Palabras clave: contenidos mentales de los otros; discurso dirigido a uno mismo; altruismo espontáneo; Teoría de la Mente; expectativas vicarias.

[1] I am very grateful to the reviewers for their helpful comments.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056
Recibido: 17th December 2021
Aceptado: 25th August 2022

1. INTRODUCTION

What in this work is understood as the most demanding type of moral capacity consists of those decisions made in favour of another person which the subject, when making them, feels are contrary to his own goals. Could any base for such type of moral capacity appear in evolution (or, in a more explicit and updated way, in the combination of "evolution in the strictly biological sense, culture and co-evolution")? My proposal tries to explain how such a thing would be possible.

When we pursue making moral and evolution compatible, the attempts that assimilate moral capacity either with "general self-control" or with "spontaneous altruism" face an easy problem, since those two senses of moral are adaptively advantageous. But regarding the most demanding moral capacity, the problem of making that capacity compatible with evolution seems much more difficult. (Note that it would also be unable to be sustained by any type of group selection. Remember the Good Samaritan.[2]) Faced with that difficulty, we might be tempted to turn to the idea that the human soul is foreign to evolution. But let us try to look for a solution.

Now I will begin by briefly differentiating the most demanding moral capacity from the abilities that appear to be near to it. Therefore, on the one hand, it will have to be separated from any type of foresighted self-control in favour of one's own interests. On the other hand, it will have to be contrasted with "spontaneous altruism".

The empirical study of self-control in favour of one's own interests began in the 1960s with the experimental paradigm of "the later but better reward".[3] Nowadays it has been strength-

---

[2] Cf. Jay ODENBAUGH, "Nothing in Ethics Makes Sense Except in the Light of Evolution?", *Synthese* 194 (2017) 1031-1055: https://doi.org/10.1007/s11229-015-0675-7: "(Internally) altruistic groups outcompete (internally) selfish groups, and, in fact, they can drive them socially extinct. (...) Thus, we found a conflict between our considered moral judgments and the implications of bio-cultural group selection."

[3] Cf. Walter MISCHEL – Ebbe EBBESEN, "Attention in Delay of Gratification", *Journal of Personality and Social Psychology* 16 (1970) 329-337: https://doi.org/10.1037/h0029815.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

ened by the current interest of the fight against addictions.[4] The self-control, traditionally called willpower, requires a special type of effort (as Baumeister has emphasised in numerous works).[5] That ability can provide many functions, amongst which we find most interesting the maintaining or improvement of the individual's relationship with his group.

That particular use of self-control has two ways. The first way, connected to an aversive motivation, is reduced to stopping the behaviours which could provoke social reprobation. The second one, which is led by hunger for prestige (that is, by a really new, peculiarly human version of the instinct of social dominance[6]), actively searches its objective. But the core of both ways is always caring for one's own reputation: The subject aspires to be recognised as someone who is to some extent important to the group (as someone who fulfils his commitments to cooperative partners).[7] More concretely, Leary argues that individuals cannot afford to wait for a threat of exclusion to be imminent, but they have to become aware beforehand of the possibilities of a decrease in relational value:[8] That would be the function of the regulatory mechanism which this author calls sociometer —"a

---

4 Cf. Peter Gollwitzer, "Weakness of the Will: Is a Quick Fix Possible?", *Motivation & Emotion* 38 (2014) 305-322: https://dx.doi.org/10.1007/s11031-014-9416-3; Angela Duckworth – James Gross, "Self-Control and Grit", *Current Directions in Psychological Science* 23 (2014) 319-325: http://dx.doi.org/10.1177/0963721414541462.

5 Cf. Roy Baumeister – Ellen Bratslavsky – Mark Muraven – Dianne Tice, "Ego Depletion: Is the Active Self a Limited Resource?", *Journal of Personality and Social Psychology* 74 (1998) 1252-1265.

6 Prestige is associated with evolutionarily new nonverbal displays: Zachary Witkower – Jessica Tracy – Joey Cheng – Joseph Henrich, "Two Signals of Social Rank. Prestige and Dominance are Associated with Distinct Nonverbal Displays", *Journal of Personality and Social Psychology* 118 (2020) 89-120: https://doi.org/10.1037/pspi0000181.

7 Cf. Nicolas Baumard – Jean-Baptiste André – Dan Sperber, "A Mutualistic Approach to Morality. The Evolution of Fairness by Partner Choice", *Behavioral and Brain Sciences* 36 (2013) 59-78: https://doi.org/10.1017/S0140525X11002202; Michael Tomasello, *A Natural History of Human Morality*, Cambridge, MA, Harvard University Press, 2016.

8 Cf. Mark Leary, "The Sociometer, Self-esteem, and the Regulation of Interpersonal Behavior", in Roy Baumeister – Kathleen D. Vohs (eds.), *Handbook of Self-regulation: Research, Theory, and Applications*, New York, NY, The Guilford Press, 2004, 373-391.

marker of one's relational value to other people". Regarding that issue, Wu *et al.* underline a necessary distinction: "While people develop a first-hand reputation from their interaction partners, they also develop a second-hand reputation through gossip".[9] Precisely nowadays there is consensus that the appearance of language —or, more specifically, of "gossip"— would have led to a more careful protection of one's own reputation.

In the past, those types of motivation were sometimes labelled as hypocritical, pharisaic, or Machiavellian ones and they acquired negative connotations. Nowadays, we know that the ability of self-control is a crucial human peculiarity, and also a decisive factor for academic and social success. But, despite all that, it must be separated from the most demanding moral capacity: In that self-control, the subject does not feel at all that his decisions are harmful for him, but, quite the contrary, they represent for him a step forward towards the non-immediate goals which were previously active in his organism.

Going on to spontaneous altruism, this concept adds two features to the traditional, philosophical one of "empathy". The first one is the concern for origins: The term "spontaneous altruism" is mainly used in the research on apes[10] and children.[11] The second feature, which derives from the first one, is an acceptance of the difference between this type of altruism and effortful decisions: That wording by Tomasello —"Why be nice? Better not think about it"— portraits that type of behaviours

---

9 Junhui WU – Daniel BALLIET – Paul VAN LANGE, "Reputation, Gossip, and Human Cooperation", *Social and Personality Psychology Compass* 10 (2016) 354 (350-564): http://dx.doi.org/10.1111/spc3.12255.

10 Cf. Frans DE WAAL, *The Age of Empathy*, Three Rivers Press, 2010; Felix WARNEKEN – Brian HARE – Alicia MELIS – Daniel HANUS – Michael TOMASELLO, "Spontaneous Altruism by Chimpanzees and Young Children", *Plos Biology* 5 (2007) 1414-1420: https://doi.org/10.1371/journal.pbio.0050184; Christopher KRUPENYE – Jingzhi TAN – Brian HARE, "Bonobos Voluntarily Hand Food to Others but Not Toys", *Proceedings of the Royal Society B* 285 (2018) 1-8: https://doi.org/10.1098/rspb.2018.1536.

11 Cf. Felix WARNEKEN, "Precocious Prosociality: Why Do Young Children Help?", *Child Development Perspectives* 9 (2015) 1-6: https://doi.org/10.1111/cdep.12101.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

well.[12] More concretely, Grossmann *et al.* show the independence between this altruism and self-control: In 18-months-old infants a "link between low latency and high frequency of prosocial behaviour exists independently of infants' ability for inhibitory control".[13]

It is very likely that the mutualism or interdependence which would have been progressively growing throughout the different hominids and the Neanderthal gave rise to the spontaneous and unthinking trend to help a peer in need.[14] Spikins *et al.*, after pointing out the unquestionable footprints of caring for the ill or the wounded that have been found in Neanderthals, highlight "the selective advantages of reducing the risk of mortality of other group members in groups whose members are highly interdependent."[15]

But we have to focus mainly on the help which, in a spontaneous and impulsive way, human adults give each other nowadays. This behaviour is carried out without any subjective effort, without any clash with one's own goals. Thus, *a fortiori*, we can say that such behaviour, just like the self-control in favour of one's own goals, dispenses with any clash with one's own main goals. In short, in spontaneous altruism there is never the feeling of losing resources or opportunities that characterises the most demanding moral capacity. In addition, let us note that that altruism does not demand from us solidarity with our neighbours who, in a stroke of luck, are placed in

---

12 Cf. Michael Tomasello, "Why Be Nice? Better Not Think About It", *Trends in Cognitive Sciences* 16 (2012) 580-581: https://doi.org/10.1016/j.tics.2012.10.006. Cf. also David Rand – Joshua Greene – Martin Nowak, "Spontaneous Giving and Calculated Greed", *Nature* 489 (2012) 427-430: http://dx.doi.org/10.1038/nature11467.

13 Tobias Grossmann – Manuela Missana – Amrisha Vaish, "Helping, Fast and Slow: Exploring Intuitive Cooperation in Early Ontogeny", *Cognition* 196 (2020): https://doi.org/10.1016/j.cognition.2019.104144.

14 Cf. Brian Hare, "Survival of the Friendliest: *Homo sapiens* Evolved via Selection for Prosociality", *Annual Review of Psychology* 68 (2017) 155-186: https://doi.org/10.1146/annurev-psych-010416-044201.

15 Penny Spikins – Andy Needham – Barry Wright – Calvin Dytham – Maurizio Gatta – Gail Hitchens, "Living to Fight Another Day: The Ecological and Evolutionary Significance of Neanderthal Healthcare", *Quaternary Science Reviews* 217 (2019) 107 (98-118): https://doi.org/10.1016/j.quascirev.2018.08.011.

a status which is higher than ours, or, in other words, here the fight against envy is not demanded. It does not include forgiveness either, which, not like animal behaviours of reconciliation, is adaptively disadvantageous, as McCoy & Shackelford —correctly, in my view— argue.[16]

Another way of understanding morality combines spontaneous altruism and foresighted self-control, and links the two of them in a developmental sequence which would appear in evolution as well as in children. I fully accept that sequence. However, I do not find the most demanding moral capacity there, not the way I understand it.[17]

In short, self-control and spontaneous altruism are both easy to explain within evolution. Indeed, these two human features can be connected with evolution in a too easy way, that is, "not by explaining how disinterested altruism has evolved, but by explaining how these behaviours are not altruistic" (as Schloss says about the connection that is offered by some proposals[18]).

---

16 Cf. Mark MCCOY – Todd SHACKELFORD, "No Such Thing as Genuine Forgiveness?", *Behavioral and Brain Sciences* 36 (2013) 28-29: https://doi.org/10.1017/S0140525X12000544.

17 Nowadays this is acknowledged in Social Decision Neuroscience studies: "Information processing can be understood as «social» or «non-social» at different levels." More concretely, processes can be prosocial at the implementational and/or the algorithmic level, while their goal is a self-benefiting one. Patricia LOCKWOOD – Matthew APPS – Steve CHANG, "Is There a 'Social' Brain? Implementations and Algorithms", *Trends in Cognitive Sciences* 24 (2020) 802-813, here 802: https://doi.org/10.1016/j.tics.2020.06.011. See also Patricia LOCKWOOD *et al.*, "Distinct Neural Representations for Prosocial and Self-benefiting Effort", *Current Biology* 32 (2022) 1-14: https://doi.org/10.1016/j.cub.2022.08.010.

18 Jeffrey P. SCHLOSS, "Darwinian Explanations of Morality: Accounting for the Normal but Not the Normative", in Hilary PUTNAM – Susan NEIMAN – Jeffrey SCHLOSS (eds.), *Understanding Moral Sentiments: Darwinian Perspectives*, Transaction, 2014, 81-121, here 91: http://dx.doi.org/10.4324/9781351296281-6. However, this author does not provide the solution that I look for, but only a lucid intensification of the problem. "Self-sacrifice is possible because we still have selves": Jeffrey SCHLOSS, "Our Shared Yearnings for a Greater Good", *Minding Nature* 10 (2017) 14-22, here 21. This is "a paradox" according to Peters (cf. Ted PETERS, "Free Will in Science, Philosophy, and Theology", *Theology and Science* 17 [2019]: https://doi.org/10.1080/14746700.2019.1596215). Or also Schloss: "There is an irony here that I find sublime. While autonomy is relinquished at the

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

However, the base for the most demanding moral capacity still today seems incompatible with evolution. That capacity seems too different from all the other capacities, and, more importantly, too disconnected from all our biological forces.

Faced with that apparent incompatibility, someone might be tempted, as said above, to turn to the idea of the human soul being foreign to evolution. But nowadays that idea is becoming implausible. Now, after leaving behind all the unmotivated taboos accepted by behaviourism, we increasingly recognize conscious cognitive abilities in many animal species.[19] Furthermore, the current focus on hominids and Neanderthals opens a new door for us which was undreamt of for previous philosophers and scholars. As a consequence of all of the above, the greatness of evolution prevails and, thus, the idea of the human soul being foreign to evolution has become more and more undesirable. Therefore, in my view, it is necessary to explain how the spiritual peculiarities of human beings could emerge in evolution. This will be an assumed premise of mine.[20] But —returning to my point— the task of explaining in such a way the base for the most demanding moral capacity involves a degree of difficulty that is not found when we try to explain, for example, mathematical or artistic creativity.

«lower» level [of Evolutionary Transitions in Individuality], it is actually enhanced at the higher" (2017, p. 16).

19 This view is really more parsimonious: cf. Arnon Lotem – Joseph Y. Halpern – Shimon Edelman – Oren Kolodny, "The Evolution of Cognitive Mechanisms in Response to Cultural Innovations", *Proceedings of the National Academy of Sciences* 114 (2017) 7915-7922: https://doi.org/10.1073/pnas.1620742114 (The typical reductionist "appeal to parsimony is somewhat misleading in evolutionary contexts and time scales, where changes are actually to be expected": p. 7916).

20 Without dualism, the promise of eternal life might seem more problematic. But, regarding that problem, let us think that God's time, unlike our poor "mental time travel", is the real simultaneity of all time (the simultaneity in which the temporal precedence, e.g., of the cause does not hold, and, in this way, the consequences of an action performed at a given point in history are present before the creation of the world). Thus, any past life, to the extent and degree to which it was somehow connected to God, is (we could suggest) already within that simultaneity, in the company of all the past and future righteous.

Now we must consider an objection. Some authors deeply rooted in the Christian spirituality seem to reject the base for the most demanding moral capacity. Those authors claim that human weakness —or more specifically, the inability to free oneself from selfish motivations— is, since it can push us towards a humble petition, the only requirement of salvation. However, other many Christian theologians underline that there is also, however weak it may be, a "natural base" for the most demanding moral capacity. Would exercising that capacity with just that resource be impossible? Of course it would be: This is why, according to Christian religion, human beings require God's Grace. But even so the "natural base" should not be despised.[21] I do not like the posture of calling for only and exclusively the supernatural help to explain the most demanding moral capacity, since that posture looks down on the whole work of evolution. In short, the existence of that "natural base" will be my second assumed premise.

In addition, of course, I bet in favour of the existence of the most demanding moral capacity. In other words, in my view, there are some human decisions that come from that capacity more than from selfish self-control or spontaneous altruism. Certainly, nowadays, there is no neuropsychological evidence regarding that question. But let us assume that that capacity, which would be the true moral freedom, does exist, and wonder: Could the base for such type of moral capacity appear in evolution (or, in a more explicit and updated way, in the combination of "evolution in the strictly biological sense, culture and co-evolution")?

All in all, we have to try to find out if and how evolution and the base for the most demanding moral capacity could be compatible. This —given the two premises that I have assumed— is a pressing question for Christians. Certainly I mean a moral capacity, not a moral doctrine. However, if that base could not emerge in evolution, then —given those premises— the

[21] There, "natural" must be understood as including "cultural" (e.g., cultural learning of language): The current, updated understanding of evolution involves "co-evolution with culture".

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

religion that preaches such doctrine would have to be considered false.

If I am allowed a biographical note, let me tell you that I read the famous statement by Ghiselin ("If the hypothesis of natural selection is both sufficient and true, it is impossible for a genuinely disinterested or altruistic behaviour pattern to evolve"[22]) many years ago, precisely when I was deciding on the topic for my dissertation for my PhD. Nowadays, that statement (and even Ghiselin's work[23]) is certainly highly discredited. It is clear that scholars at that time were not aware of the broadness of spontaneous altruism. But, in spite of all such limitations, I continue to believe that the challenge proposed by Ghiselin contains a nucleus to which we must respond. Indeed, I was greatly affected when I read his statement. More concretely, with regards to my dissertation, it led me to lean towards a topic which was —as I saw it— difficult and daring.

In my dissertation,[24] I proposed that predicative language, and, consequently, syntax originate (both in children and in evolution) when an individual —the future speaker of predication— grasps that the meaning which another individual has with regards of a particular reality is wrong or incomplete or not up to date. Such wrong or incomplete meaning would be the *theme*, which the speaker would correct or complete by adding the *rheme* ("*theme / rheme*" is most likely the original type of syntax). Thus, the decentering which is of interest, that which is particularly human, would not be the visual-spatial one on which Piaget focused (the Three Mountains test, or, even easier, the 6 which becomes a 9 for that who is looking at it across the table), but the grasping of that wrong or outdated meaning. In the dissertation, I mainly focused on Frege and Piaget. But in 1988, I discovered that what I had called "the addressee's meaning" was being studied (under the name of

---

22 Cf. Michael T. Ghiselin, "Darwin and Evolutionary Psychology: Darwin initiated a radically new way of studying behavior", *Science* 179 (1973) 964-968: https://doi.org/10.1126/science.179.4077.964.

23 Cf. also Benjamin J. Fraser, *Sexual Selection and the Evolution of Morality.* (PhD Thesis), Camberra, Australian National University, 2010, 46-48.

24 Cf. Teresa Bejarano, *Comunicación descentrada y creatividad* (Thesis PhD), Universidad de Sevilla (no publicada), 1985.

"false belief") by the so-called "theory-of-mind" (ToM). All that line of my research was collected much later.[25]

But let us return to what is of interest here. The work involved in writing my dissertation showed me that the grasping of another's interiority constitutes the basis for, at least, some of the most important human particularities. For instance, syntax (which, according to my proposal, was absent from pre-linguistic thought and derived from that grasping[26]) transformed thought crucially.[27]

Therefore, we should ask ourselves how that grasping could be related to the possibility of "loving your neighbour as yourself". I began by elaborating on the difference between such love, which in the decisions of the most demanding moral capacity becomes so demanding and distressful, and non-demanding altruisms or even "the pleasant tears with fiction".[28]

---

25 Cf. Teresa BEJARANO, *Becoming Human: From pointing gestures to syntax* (Advances in Consciousness Research 81), Amsterdam, Benjamins, 2011.

26 Certainly Jerry FODOR, *The Language of Thought*, Cambridge, MS, Harvard University Press, 1975, postulated a "syntactic, innate language of thought", which would go with all perceptions. However, I lean towards rejecting it. In linguistic reception, each of the meanings receives independent attention before they are integrated into the total meaning. However, in perceptions, such attentional (non-subpersonal) independence of each relevant element would be a detrimental feature. Thus, syntax originated —I propose— dialogically, by interactive users of non-syntactic "words": Note that "holophrastic" (that is, single-"word") calls or petitions could reveal their speaker's false belief (Cf. Teresa BEJARANO, "From Holophrase to Syntax: Intonation and the Victory of Voice over Gesture", *Humana.mente: Journal of Philosophical Studies* 27 [2014] 21-37). This would be the evolutionary emergence of what Moore calls "the conversational experience of clashing perspectives" (Richard MOORE, "The Cultural Evolution of Mind-Modelling", *Synthese* 199 [2021] 1751-1776: https://doi.org/10.1007/s11229-020-02853-3).

27 Cf. some suggestions in Teresa BEJARANO, "Un aspecto de la relación entre pensamiento y lenguaje. De la interrogación interpersonal a la resolución creativa de problemas", en H. VAN DITMARSCH – F.J. SALGUERÓ – F.SOLER (eds.), *Liber Amicorum Ángel Nepomuceno*, Sevilla, Fénix, 2010, 27-34; and Teresa BEJARANO, "Buscando el requisito decisivo para el origen del número", en C. BARÉS – F.J. SALGUERO – F. SOLER (eds.), *Lógica, Conocimiento y Abducción*. Homenaje a Ángel Nepomuceno, London, College Publications, 2021, 131-152.

28 Cf. Teresa BEJARANO, "Las emociones ante la ficción", *Thémata* 13 (1995) 73-95; IDEM, "Libertad moral y evolución biológica: un interrogante

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

However, it's only now when I have come to develop a proposal. But enough with my biographical note.

Many answers have been given to Ghiselin's challenge. It has often been proposed that an adaptively disadvantageous behaviour may be (this is called "pleiotropy") associated with a by-product of other phenotypes that are adaptive. An example which is usually offered is human or non-human behaviour of adoption, which can be viewed as being associated with genes for bonding with one's progeny. But that bonding is so crucial that its involvement in those (very infrequent) circumstances does not question its adaptive advantage. Besides, human adoption, at least in our society, is a strong goal in the subject, while the most demanding moral decisions are opposed to the previously strongest goals.

Ayala certainly uses "pleiotropy" in a way closer to my purposes.[29] In his words, human morality is a pleiotropic consequence of reason and self-awareness, both of which are adaptations. However, that statement overlooks the difference between moral and non-moral uses of reason, or the relation between self-awareness and awareness of somebody else's mental contents.

I propose that the base that we are looking for —that is, the natural base for the most demanding moral capacity—, while it is not an evolutionarily adaptive feature but rather the opposite, arises, however, from the evolutionary convergence between two abilities —the advanced mode of ToM, and mature inner speech—, which were extremely advantageous in their respective origins.[30] (I develop this proposal more widely in a book in progress).

---

que conviene ya ir planteando", *Thémata* 34 (2005) 235-247; and Idem, "Autorregulación y libertad", *Thémata* 43 (2010) 65-86.

29 Cf. Francisco Ayala,"The Difference of Being Human: Ethical behavior as an evolutionary by-product", in H. Rolston (ed.), *Biology, Ethics, and the Origins of Life*, Boston, MS, Jones & Bartlett, 1995, 113-136.

30 Antonio Benítez-Burraco – Francesco Ferretti – Ljiljana Progovac say about language: "A trait as baroque and bizarre from the point of view of nature could evolve thanks to an exceptional convergence of factors" ("Human Self-Domestication and the Evolution of Pragmatics", *Cognitive Science* 45 [2021]: https://doi.org/10.1111/cogs.12987). This can even more strongly be said about that moral capacity.

2. THE TWO MODES OF ToM

We have to elaborate the difference between the primitive mode and the advanced mode of ToM. But let us very briefly attend, first of all, to what has been called ToM. In 1978 the ability to infer somebody else's mental elements becomes an important issue, and is named "theory-of-mind". It is discovered that children up to 4 years of age are unable to perform effectively on "false belief" tests. Those tests show a video in which, for example, a child (Maxi) puts his marble inside a vase and then leaves; afterwards, his mother puts the marble inside his toy box and leaves. Right then, Maxi comes back and the experimenter asks the children who have seen the video, "Where will Maxi look for his marble?" The answers coming from children under 4 do not show the false belief which Maxi is bound to have, but their own knowledge.

Nowadays, as said above, the main focus is on the two modes —the primitive and the advanced— of ToM. Let us see how these were originally described. The advanced, uniquely human mode was identified in verbal tests of someone else's false beliefs —Maxi's test, for example—, which require the ability to distinguish and compare reality and others' beliefs. (In principle, for any subject, his current belief is knowledge — is reality.) On the other hand, the primitive mode would include abilities of children under 4, and also of chimpanzees. Tomasello *et al.* showed that a chimpanzee can estimate what other —the dominant individual, for example— sees or has seen from the specific location in which he —the dominant one— is placed. Thus, the subordinated individual typically will (/will not) catch a piece of food if the dominant one is (/is not) ignorant of the existence of the piece.[31]

But after Onishi & Baillargeon, non-verbal tests of false belief have been used, which were successfully passed by

[31] Cf. Michael TOMASELLO – Josep CALL– Brian HARE, "Chimpanzees Understand Psychological States — the Question is Which Ones and to What Extent", *Trends in Cognitive Sciences* 7 (2003) 153-156: https://doi.org/10.1016/s1364-6613(03)00035-4.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

infants in numerous experiments, and also by chimpanzees.[32] However, there is much controversy regarding those results. The only clear thing is that, for both chimpanzees and infants, those tests are more difficult than the attribution of ignorance.[33]

However, whether infants and chimpanzees infer or not the "false belief" of others, I accept that human adults have two different modes to understand somebody else's mental states. Apperly & Butterfill were the first to point to a primitive mode in adults.[34] But now we should go over how the primitive mode has been recently re-described.

According to Tomasello,[35] in the primitive mode, not like in the advanced one, the infant grasps others' beliefs because he "disregards his own (diverging) knowledge". Such formulation is not convincing, since disregarding the knowledge of the situation in which we find ourselves would be a very inconvenient type of inattention. But it is also true that, as Tomasello argues, if one's own mental content, instead of being disregarded, is simultaneously carried with somebody else's content in one's own mind, then the two contents must be distinguished and compared by the subject, and thus, we would be identifying the primitive mode with the advanced one — an identification that I reject.

---

32 Cf. Kristine Onishi – Renée Baillargeon, "Do 15-Month-Old Infants Understand False Beliefs?", *Science* 308 (2005) 255-258: https://dx.doi.org/10.1126%2Fscience.1107621; Fumihiro Kano – Josep Call – Christopher Krupenye, "Primates Pass Dynamically Social Anticipatory-Looking False-Belief Tests", *Trends in Cognitive Sciences* 24 (2020) 77-778: https://doi.org/10.1016/j.tics.2020.07.003.

33 Why are they more difficult? In order to pass tests of false belief, two different scenes must be attended to jointly in working-memory. Since developmentally —and likely also evolutionarily— a great expansion of working memory is caused by the reception of multiple-word messages, could we explain in this way that greater difficulty? I can't answer this question, but I will propose that the essential, founding difference between the two modes of ToM is not "ignorance vs. false belief", but "vicarious expectation vs. foreign mental content".

34 Cf. Ian A. Apperly – Stephen A Butterfill, "Do Humans Have Two Systems to Track Beliefs and Belief-Like States?", *Psychological Review* 116 (2009) 953-970: https://doi.org/10.1037/a0016923.

35 Michael Tomasello, "How Children Come to Understand False Beliefs: A Shared Intentionality Account", *Proceedings of the National Academy of Sciences* 115 (2018) 8491-8498: https://doi.org/10.1073/pnas.1804761115.

Thus, I agree with Tomasello that the union of "inattention to one's own mental elements" and "attention to somebody else's ones" characterises the primitive mode. But I propose that such inattention and such attention take place, not at the content-level, but at the expectation-level. It is the subject's expectation that is disregarded or 'deactivated' in (non-human or human) primitive mode.

But, before developing my proposal, we must focus on another account, which, being relatively similar to Tomasello, is more recent and elaborate. Southgate proposes that human infants have an altercentric bias ("which results from a combination of the value that human cognition places on others, and an absence of a competing self-perspective"), and that such bias causes that the particular events (for example, the displacement of Maxi's marble) that are not co-witnessed with the protagonist of the play are encoded with less strength. This is what explains, according to Southgate, infants' successes in non-verbal tests of false belief.[36]

I will start by saying that I like the idea that "for infants, altercentrism is beneficial".[37] But that tendency towards altercentrism can, without losing any of its attractiveness, be reformulated by saying that the infant very often produces "vicarious expectations" (we will focus on these right away), which mostly correspond to the circumstances of his carers.[38] In addition, I

---

[36] Cf. Victoria SOUTHGATE, "Are Infants Altercentric? The Other and the Self in Early Social Cognition", *Psychological Review* 127 (2020) 505-523: https://doi.org/10.1037/rev0000182.

[37] Southgate cites Stein BRÄTEN, "Hominin Infant Decentration Hypothesis: Mirror neurons system adapted to subserve mother-centered participation", *Behavioral and Brain Sciences* 27 (2004) 508-509: https://doi.org/10.1017/S0140525X0427011X. See also Vittorio GALLESE, "The Problem of Images: A view from the brain-body", *Phenomenology and Mind* 14 (2018) 70-79: https://doi.org/10.13128/Phe_Mi-23626.

[38] The same can be said of the results by Charlotte GROSSE WIESMANN – Angela D. FRIEDERICI – Tania SINGER – Nikolaus STEINBEIS, "Two Systems for Thinking about Others' Thoughts in the Developing Brain", *Proceedings of the National Academy of Sciences* 117 (2020) 6928-6935: https://doi.org/10.1073/pnas.1916725117, who show that "the network that supports nonverbal ToM reasoning" is also "involved in emotional and visual perspective-taking, and action observation" (p. 6928): Those results fit with Southgate's hypothesis, but also with what I will propose.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

do not need to call on the alleged "weakness of self-perspective": According to my proposal, neither one's own knowledge nor one's own perception are weakened or unattended in the primitive mode of ToM. Note that typical perceptions are evolutionarily older than the altercentric ones and are used at any age much more frequently. Thus, it is unlikely that the degree of conservatism that evolution necessarily includes fails there.

Moving on to my proposal, I shall start by underlining that animals are goal-driven and tend at all times to activate expectations (well-defined voids, so to speak) not only of homeostatic goals, even before experiencing them for the first time, but also of learned associations. Expectation guides perception and recognition. It is also the resource par excellence which animal learning uses.

Later, a special type of expectations appeared. These special, vicarious expectations correspond (in the manner of what happens in mirror-neurons) to the circumstances of the individual observed by the subject, and they (e.g. reward prediction errors) "are encoded as «belonging to other» since the very beginning of their computation".[39] But, despite all this, vicarious expectations, like any other expectation, have to always be intrinsically possible for the subject (i.e., possible at some time and circumstances different from his current situation). Otherwise, the subject would not possess the corresponding "well-defined void".

Having proposed that in the primitive mode of ToM the subject pays attention to expectations that correspond to the circumstances of the observed individuals and disregards the expectations which correspond to his own circumstances, I must add that this disregarding —this inattention— can only take place when the subject is behaviourally inactive. Please note that in the middle of a behavioural pattern, the subject needs to focus on the expectations which correspond to his own circumstances. In short, the primitive mode can only appear in the following way: behavioural inactivity > possibility

[39] Sam Ereira – Raymond J. Dolan – Zeb Kurth-Nelson, "Agent-specific Learning Signals for Self-other Distinction During Mentalising", *Plos Biology* 16 (2018): https://doi.org/10.1371/journal.pbio.2004752.

of disregarding one's own expectations > vicarious expectations > the primitive mode of ToM.

Now, in order to enquire into the origin of the advanced mode —the origin of "foreign" mental contents—, we must ask ourselves: For what task were vicarious expectations insufficient for the first time? I assume the following three points. First, in order to support the primitive mode, vicarious expectations are sufficient. Second, the nature of any kind of expectation implies that a subject can only have expectations of states which are intrinsically possible for him, for the subject. Third, the state of interacting with the subject as with a different person is not an intrinsically possible state for the subject. Therefore, the bankruptcy of vicarious expectations and the emergence of the advanced mode of ToM probably arose when the ability of grasping foreign mental states that involve oneself became an advantageous one.

Thus, I propose that "the thinking what others think *of us*" (Darwin, in relation to blush[40]) necessarily requires the advanced mode of ToM. But, beyond blush, we can focus on self-conscious emotions: embarrassment, pride, shame, and guilt.[41] These emotions are adaptively advantageous.[42] Therefore, the early "advanced mode" would be connected with them.

However, the process of thinking of foreign mental states which involve oneself —that particular process— would be a requirement only *for the origin* of the advanced (probably uniquely human) mode. In fact, I propose that, once the ability to think a "second line" of contents was reached, the advanced mode can carry complex functions which do not fulfil that requirement. Such functions sometimes use foreign but non-interactive contents, as in verbal Maxi-test, which is (I borrow Dor's words) "an individualistic, non-dialogic capacity

---

40 Charles DARWIN, *The Expression of the Emotions in Man and Animals*, London, John Murray, 1872 (chapter 13, pp. 326-327, my emphasis).

41 Cf. Michael LEWIS, "The Emergence of Human Emotions", in Michael LEWIS – Jeannette HAVILAND-JONES (eds.), *Handbook of Emotions*, New York, NY, Guilford Press, 2000, 265-280.

42 Cf. LEARY, "The Sociometer", or Daniel SZNYCER, "Forms and Functions of the Self-Conscious Emotions", *Trends in Cognitive Sciences* 23 (2019) 143-157: https://doi.org/10.1016/j.tics.2018.11.007.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

of mind-reading".[43] Other times, they use non-foreign (e.g., "possible" in the sense of "maybes") contents. Nevertheless, originally the advanced, probably uniquely human mode of ToM is —let us say it again— a directly relational, interpersonal process.

I wonder[44] if the ability of grasping somebody else's mental contents could perhaps originate in the specifically human reception of pointing gestures — a more basic, more immediately useful ability than self-conscious emotions. But, one way or another, we can reach the desired conclusion. The very origin of the advanced mode of ToM was linked to an adaptive advantage.

3. THE TWO MODES OF ToM AND THE DIFFERENCE BETWEEN SPONTANEOUS ALTRUISM AND THE MOST DEMANDING MORAL CAPACITY

The difference between spontaneous altruism and the base for the most demanding moral capacity is not a merely quantitative one. The crucial evolutionary novelty that the advanced mode of ToM meant —that is, the grasping of somebody else's mental contents— was necessary in order for that base to be provided. This is what I will try to show now.

I propose that spontaneous altruism is a very sporadic derivation of the primitive mode of ToM.[45] Regarding this mode, I have proposed, firstly, that it relies on expectations that correspond to another individual's circumstances, and, secondly,

---

43 Cf. Daniel Dor, "From Experience to Imagination: Language and its evolution as a social communication technology", *Journal of Neurolinguistics* 43 (2017) 107-119: https://doi.org/10.1016%2Fj.jneuroling.2016.10.003.

44 Cf. Teresa Bejarano, work in progress.

45 My understanding of any spontaneous altruism might be near to a recent proposal about infants' altruism: "The infant [in his «goal slippage»] identifies the goal of the observed individual in the lean (*vs.* rich) sense of an outcome —event— toward which the agent's movements are directed": (cf. John Michael – Marcell Székely, "Goal Slippage: A Mechanism for Spontaneous Instrumental Helping in Infancy?", *Topoi* 38 [2019] 173-183, here 180: https://doi.org/10.1007/s11245-017-9485-5). That lean (*vs.* rich) sense may correspond to vicarious expectations (*vs.* full foreign contents).

that those special, vicarious expectations can arise only when the subject is inactive. Such inactivity can sometimes be due to the subject needing to collect a particular piece of information before acting,[46] but other times —and this second type is what interests us now— inactivity can be due to the subject having no goal urgently activated.

Spontaneous altruism can, in my view, arise only in the moments of that second type of inactivity, i.e. only in those moments in which the subject, in an attitude such as that of a spectator and without being driven by any goal, looks at others and has the corresponding vicarious expectations.[47] This is why the activation of spontaneous altruism mainly depends on the subject's previous state —or, more concretely, on his / its "spectatorial" attitude— and not on the state of the other individual. In addition —since in spontaneous altruism the subject is following, not foreign contents, but vicarious expectations, which belong to the range of mental elements which are intrinsically possible for him— that type of altruism can immediately and directly find motivational force. However, in the most demanding moral capacity none of those two points appears. (Spontaneous altruism and the most demanding moral capacity are respectively

---

[46] Cf. again TOMASELLO – CALL – HARE, "Chimpanzees Understand Psychological States", 153-156.

[47] Let us specify the connection between spontaneous altruism and the attitude of the spectator of fictions. The prosperity of the fictional character with which we have identified ourselves brings us joy. However, if it is our neighbours who, in a stroke of luck, are placed in a status which is higher than ours, then, feeling solidarity can require a considerable effort. (That difference between fiction and reality appears, however, in a much lesser degree in the case of suffering. Indeed, unless helping behaviours which clash with the subject's goals are required, the observation of others' pain in the real world is, except in some psychopaths, as empathic as the reaction of the viewer of fiction). Why, when facing the prosperity of those who share our environment, does empathy disappear? This limitation ultimately depends, of course, on the evolutionary origin of empathy, that is, on the increasing need for a cooperative lifestyle (let us remember Neanderthals). But its "proximate cause" lies in the observed prosperity becoming a goal for the observer. As said above, when one's own goals are activated, vicarious expectations that aren't sub-goals regarding those goals disappear, and, consequently, also spontaneous altruism does.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

invoked by the first part and the second one of some parables — since Nathan's parable to The Prodigal son).[48]

Let us move on to the connection between the advanced, uniquely human mode of ToM and the base for the most demanding moral capacity. That mode is able, as said above, to compare one's own and others' mental contents. Now, we must underline that the grasping of foreign contents includes foreign needs, and these needs incorporate the assessment mechanism that belongs to the other individual: In other words, the urgency grasped in those needs typically matches the urgency which they truly have for the other individual. Thus, there are now two different assessment mechanisms in the subject's mind (which was to be expected, since there are one's own and somebody else's contents). Now the subject understands the convenience of actions which he knows would be contrary to his own interests. That is a great novelty.

Nevertheless, it must be immediately added that such novelty is still unable to offer even the poor and weak base that we are searching — the "natural" base for the most demanding moral capacity. Certainly, the urgency involved in those foreign states can be very strong, and that urgency is grasped by the observer: that is the side that we have just seen. However, because they are states of the observed individual, that urgency depends only on the assessment mechanism that belongs to the observed individual. Thus, the other assessment mechanism —the one that belongs to the observer subject— will issue a different decision and, therefore, that grasping of urgency will have to *face the goals* that (not like in spontaneous altruism, where the subject has *no active goal)* are active in the subject at that time, which are the ones which count on sufficient motivational force.[49]

But let us keep focusing on what happens when the grasping of foreign contents has to face the goals which are

---

48 Cf. Teresa Bejarano, "Parábolas, altruismo espontáneo y coherencia cognitiva. Analizando la eficaz construcción de algunas parábolas", *Isidorianum* 29 (2020) 13-36: https://doi.org/10.46543/ISID.2029.1053.

49 Beyond moral decisions and spontaneous altruism, the third element — selfish self-control— *moves towards goals*, which, despite lacking immediacy, are active beforehand.

active in the subject at that moment. We have, on the one hand, the fact that the idea of undertaking the behaviours of very costly help does not receive any positive motivation derived from the subject's interests, but quite the opposite, a considerable negative motivation. However, on the other hand, that grasping would have properly detected the foreign need and its degree of urgency: That information is true.

Note that what I have just called true is precisely the information which the subject has grasped regarding another individual's needs and interests. Therefore, when I speak of such truth, I am not referring to the truth or untruth of moral judgements. The latter, however, concerns many authors, and it divides them: Thus, for example, the scepticism regarding that truth argues that the world simply isn't furnished with the properties and relations necessary to render such judgements true.[50] But I'm not sharing that priority attention to moral judgements, which is common both to sceptics as well as to their adversaries. It is the process of decision-making that is the focus of this work. Thus, the nucleus of the novelty involved in the base for the most demanding moral capacity has to do with the fact that in the advanced mode of ToM the two typical features of perceptions —one, that of informing about the surroundings, and the other, that of being useful to the subject's interests— would be, for the first time in evolution, dissociated from each other.

Returning to our thread, the human subject can understand that the foreign needs are as real and as urgent as their own. That knowledge is certainly a jewel amongst cognitive abilities.[51] However, it is likely that the behavioural

---

50 Richard JOYCE, *The Evolution of Morality*, Cambridge, MA, MIT Press, 2007, argues that natural selection is a "belief pill" for certain moral beliefs about cooperation."Thus, although these might actually be morally true, knowledge that your belief is the product of a belief pill renders the belief unjustified".

51 The ability "to generate representations of intentional relations that are uniformly applicable to the activities of both self and other" (John BARRESI – Chris MOORE, "Intentional Relations and Social Understanding", *Behavioral and Brain Sciences* 19 [1996] 107-122: https://doi.org/10.1017/S0140525X00041790) is certainly related to the one that I focus on. However it is too general.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

recommendations which that foreign mental content entails are contrary to the active goals in the subject at that time, and, consequently, those recommendations (and even the urge to keep listening, of looking at, the observed individual) are a losing goal in the assessment mechanism that belongs to the subject. Of course, there will be a certain degree of competition. But the farthest that that situation is likely to get —the maximum result which the objective and impartial knowledge could aim to— is a brief favourable fluctuation which, in addition to disappearing right away, lacks any strength, and is therefore unable to be the natural base for the most demanding moral capacity.

4. INNER SPEECH

After the immediately previous lines, we can glimpse where inner speech could be placed within this frame, or, more specifically, the inner speech that is directed to oneself and completely mature. This type of speech —I propose— is able to take advantage of that very brief favourable fluctuation. Firstly, that speech is an extremely cheap resource and, in addition, the episode of inner speech which would be useful here could be very short (reduced, for example, to name the person whose needs we know[52]). Secondly, it is able to exercise a strong influence: More specifically, with it, the subject manipulates his own attention, and he can therefore try to change the previous relationship between motivational forces. (Certainly, these two traits —little cost and great influence— are more accentuated in inner speech. However, they also occur in interpersonal speech).

But before really including inner speech in my proposal, I must distinguish its two functional types: the speech directed

[52] This example means that I reject the statements that the mature inner speech has a dialogical structure and is immediately analogous to the second-person standpoint (cf. Charles Fernyhough, "The Dialogic Mind: A dialogic approach to the higher mental functions", *New Ideas in Psychology* 14 [1996] 47-62: https://doi.org/10.1016/0732-118X(95)00024-B). Contrary to those views, I underline the huge novelty —the creative "reuse"— which meant putting linguistic resources at the service of the task of manipulating one's own attention.

to nobody or, in other words, of emotional discharge, to which unfortunately Vygotsky's examples were reduced to,[53] and the speech that the subject directs to himself, which is the only type which interests us here. Regarding "speech directed to oneself", a question appears in many authors: what can a speaker communicate to himself? It is clear that its function cannot be the transmission of new information: Nothing can be new for a hearer who is also the speaker. However, the manipulation of attention is a function which, present in interpersonal language, is perfectly able to be intra-personalised. I propose that the function of having an influence on one's own attention is preferable to the too specialised and restricted function ("of making commitments to oneself") which Geurts has assigned to inner speech.[54] In short, in my view, the manipulation of one's own attention is the only immediate objective of speech directed to oneself. That idea (the manipulation of attention being executed by the speaker on himself[55]) can be already found in Augustine of Hippo: "When we pray, words are not necessary to inform or put pressure on God about our true needs, but to keep these needs in our mind, and, in this way, to exercise in us that desire by which we may receive what He prepares to bestow".[56]

We must also focus on the different stages of speech directed to oneself. From a developmental point of view, Vygotsky proposed that inner speech derives from external

---

53 Cf. Lev VYGOTSKY, *Thought and Language*, MIT Press, 1962 (1934).

54 Cf. Bart GEURTS, "Communication as Commitment Sharing: Speech acts, implicatures, common ground", *Theoretical Linguistics* 45 (2019) 1-30: https://doi.org/10.1515/tl-2019-0001. In Bart GEURTS ("Making Sense of Self Talk", *Review of Philosophy and Psychology* 9 [2018] 271-285: https://doi.org/10.1007/s13164-017-0375-y), the function of renewing or reaffirming a previous idea is admitted ("Don's promising himself to mow the lawn might be a way of renewing or reaffirming an intention predating his speech act": p. 281 note 7). However, this author does not generalise that function of renewing (renewing in one's own attention) as the key for all speech directed to oneself.

55 Andy CLARK,"Material Symbols", *Philosophical Psychology* 19 (2006)1-17.

56 Saint AUGUSTINE, *Letter 130 (to Proba)*, 8. Prayer and spiritual meditation involve (besides dialogue) speech directed to oneself, but they do not have to be at all brief or without motor unfolding, since they are special moral decisions that remotely prepare the moral decisions-in-action.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

speech according to a gradual process of internalization, with children under 6 or 7 years only able to "think out loud". Certainly this claim has been attacked: Mani & Plunkett allegedly showed that 18-months-old infants can implicitly name objects.[57] However, these experimental results are most likely caused by infants' expectations of receiving the name. (Gambi *et al.*: "Graded prediction ability may support linguistic development by increasing the fluency with which children process language".[58])

The sequence of stages[59] leads to the fully mature form, in which inner speech gets to have, in my view, an almost nil cost, thanks to the fact that there, the sequential (originally motor, articulatory-phonetic) format of each word is omitted or "abstracted". It is true that —as concluded by Perrone-Bertolotti *et al.*[60]— for this hypothesis, which has been called "abstraction", there is mixed evidence. However, I would say that the experimental designs created to study such abstraction are mostly focused on episodes of inner speech which are of no use or utility for the subject, except that of fulfilling the instruction — that of producing silently a certain word or syllable or tongue-twister.

---

57 Cf. Nivedita Mani – Kim Plunkett, "In the Infant's Mind's Ear: Evidence for implicit naming in 18-month-olds", *Psychological Science* 21 (2010) 908-913: https://doi.org/10.1177/0956797610373371.

58 Chiara Gambi – Priya Jindal – Sophie Sharpe – Martin Pickering – Hugh Rabagliati, "The Relation Between Preschoolers' Vocabulary Development and Their Ability to Predict and Recognize Words", *Child Development* 92 (2021) 1048-1066: https://doi.org/10.1111/cdev.13465.

59 Cf. Alexander R. Luria, *Language and Cognition*. New York, John Wiley & Sons, 1982 (orig. 1979) chapter 6; Fernyhough, "The Dialogic Mind"; and Takashi Hanakawa, "Organizing Motor Imageries", *Neuroscience Research* 104 (2016) 56-63: https://doi.org/10.1016/j.neures.2015.11.003. In Hanakawa, see mainly the sub-sub-type that this author calls "motor imagery" of "planning stage" and "of an implicit type (that is, without any sensorial evocation)". That sub-sub-type is really minimally motor, and could be close to the proposal (the final disengagement of fully mature inner speech from the motor-sequential format) which I defend.

60 Cf. Marcela Perrone-Bertolotti – L. Rapin – J. P.Lachaux – M. Baciu – H. Lœvenbruck, "What Is That Little Voice Inside My Head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring", *Behavioural Brain Research* 261 (2014) 220-239: https://doi.org/10.1016/j.bbr.2013.12.034.

That explains that those experiments are not able to show the characteristics of functional, fully mature speech.[61]

But how did the speech directed to oneself evolutionarily arise? Let us start by looking at some laboratory tasks. For example, there are toy lorries and boats which may be red or blue, and that also, with no correlation with the classifications above, are sometimes made of plastic and other times are made of wood. The subject is first asked to choose lorries, and, after some choices, he is asked to choose blue toys, and, again, after a few choices, he is given a new criterion. Perseverations following the previous criterion are often produced in that task. But if the child is asked to repeat to himself in each trial the instructions received (or merely their main word, which implicitly includes its recent links), the results improve.[62] With regard to adults, they can take the initiative of repeating the instructions.

But considering this difficult type of tasks —so an objector could say—, we cannot at all take for granted that it was there in the evolutionary origin of "speech directed to oneself". I completely disagree with that. I propose that since there were linguistic petitions (or even only pre-syntactic petitions) we have a highly special situation in the recipient. Note that the producer of the petition lies on the animal ability to pay attention to objects connected to the action in course. This is why, in order to produce the petition, no special effort is required. However, the recipient of the petition will have to search — and the search may be long— attending to a criterion which is not imposed by his previous behavioural framework. This type of criterion is a difficulty found up to that point in evolution. More concretely, these petitions —which have to replace

---

61 More in general, the proposed "abstraction" fits well with the results of Xiuyi WANG *et al.* ("Physical Distance to Sensory-motor Landmarks Predicts Language Function", *Cerebral Cortex* [2022] 1-14: https://doi.org/10.1093/cercor/bhac344), who show that "the language functions of cortical regions are related to their physical distance to sensory-motor cortex", and, in this way, suggest that separation from sensory-motor codes would be required for higher cognition.

62 Cf. Sabine DOEBEL – Philip ZELAZO, "Bottom-up and Top-down Dynamics in Young Children's Executive Function: Labels aid 3-year-olds' performance on the Dimensional Change Card Sort", *Cognitive Development* 28 (2013) 222-232: https://doi.org/10.1016/j.cogdev.2012.12.001.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

"imperative pointing gestures" when asking for objects which are far— turn to be extremely demanding for their recipient: It is difficult for him to remember them until he can satisfy them. In short, language raised the demand on the executive function of the recipient of petitions. But if language was responsible for that difficulty, then so it was for the remedy as well. Thus, a new linguistic resource appeared, with which the recipient can repaint those petitions in his own attention. This —I propose— may have been the origin of any speech directed to oneself. Please, note that in the human, cooperative style of life such origin would have been adaptively very advantageous.

Later, this speech turned out to be a very useful resource for many adaptively advantageous purposes that required more complex self-control (e.g. human work, deliberate practice in order to acquire any expertise, and —as an instrumental resource towards those purposes— manipulation of timeframe).[63] It is not surprising, therefore, that it was object for subsequent transformations (internalisation and "abstraction of motor sequence") which made it less costly, and, in this way, more able to sustain increasingly complex self-control. In short, it is extremely likely that the original "non-silent speech directed to oneself" as well as its later transformations were in their respective origins adaptively advantageous resources.

Let us return to the role that inner speech fulfils in order to build the base for the most demanding moral capacity. As above said, perceptions serve to the subject's interests and represent the surroundings adequately, and those two aspects of the perceptions were inseparable until the ability to grasp (to "see in the broad sense") "foreign" mental contents appeared. From that moment on, both aspects can clash with

---

[63] James Green – Penny Spikins, "Not Just a Virtue: The evolution of self-control", *Time & Mind* 13 (2020): https://doi.org/10.1080/1751696X.2020.1747246, focus on the evolutionarily earliest uses of self-control. Matt J. Rossano, "Ritual as Resource Management", *Philosophical Transactions Royal Society B* 375 (2020): http://dx.doi.org/10.1098/rstb.2019.0429 (an article about collective rituals that enhance psychological states) says that "psychological states can be understood as «resources», not unlike material resources such as food", and also that "psychological states are often subject to fatigue". Here, regarding, not collective rituals, but "speech directed to oneself", I repeat those claims.

one another. More specifically, the condition of true information of a grasped "foreign" need will cause some fluctuation favourable to helping behaviour, although that fluctuation, since the organism's goals are imperative, cannot be but brief, weak and destined to disappear. But if there, the subject (despite the pain that the conflict with his other goals causes in him) decides to insert a resource of such a low cost and such high performance as an episode of mature inner speech, then the favourable fluctuation could become a little —just a little— less weak.

5. CO-EVOLUTION

Now, after having focused on inner speech, is the time to comment the updated definition of evolution that was mentioned above. According to the previous sub-proposal, inner speech (dependent on "the cultural learning of language", which was, of course, sustained by cognitive abilities) provokes a new cognitive ability: the complex self-control. More in general, language both in History and in ontogenesis triggers new cognitive abilities, and thus, a double causal direction can be found between language and cognition.

The double direction between cognition and language (or "the cognitive bonus of language"[64]) was already proposed by Vygotsky.[65] According to this author, language —both when it is used interpersonally and when, later, it becomes inner speech— is able to strengthen thought, and, more specifically, to make a type of thought which exceeds that which had been a previous requirement for the emergence of language. As the reader already knows, I do not agree with the details of that Vygotskian proposal. However, I consider the very general principle to be highly valuable. (On the one hand, in my view —as said in note 26— syntax and *independent* concepts are absent from prelinguistic thought. On the other hand, we have just seen the consequences of mature inner speech).

---

64 Cf. Andy CLARK, "Language, Embodiment, and the Cognitive Niche", *Trends in Cognitive Sciences* 10 (2006) 370-374: https://doi.org/10.1016/j.tics.2006.06.012.

65 Cf. VYGOTSKY, *Thought and Language.*

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

That "double causal direction" between language and thought, or between culture and cognition, performs at two different levels (or time scales). On the one hand, it operates within the structure of the individual brain ("neural reuse"). The human brain is peculiarly plastic: There is complete evidence that its physiological, functional, and structural features can be modified by experience and practice. On the other hand, that double causal direction operates in evolution.

About this latter level, there are two theories which currently face each other. One of those theories is the S(tandard) E(volutionary) T(heory), which stays faithful to the old pair "random genetic variation and natural selection". According to this view, culture is regarded as an effective causal factor in human evolution, but only in the sense that cultural environments exert specific selective pressures on organisms. The other theory —the E(xtended) E(volutionary) S(ynthesis)—, going beyond the old pair, proposes new causes of evolution. Developmental bias, plasticity, niche construction, and extra-genetic inheritance: For SET, these phenomena are just outcomes of evolution. For EES, they are also causes.[66] Likewise, Eva Jablonka *et al.* sustain that the key to human evolution is the fact that "we constructed the cultural niche collectively, phenotypically accommodated to it individually, and eventually became, through genetic accommodation, even better niche constructors and phenotypic accommodators".[67]

But let us move on to a more concrete matter. Colagè & D'Errico defend that "cultural innovation may have effects on the evolution of cognitive capabilities of populations without the need of genetic changes".[68] In this way these authors offer "a general mechanism for cognitive evolution in which culture is the driving force", or, in other words, "a top-down

---

66 Cf. Kevin Laland *et al.*, "Does Evolutionary Theory Need a Rethink? Yes, urgently", *Nature* 514 (2014) 161-164.

67 Eva Jablonka – Simona Ginsburg – Daniel Dor, "Cognitive Gadgets and Genetic Accommodation", *Behavioral and Brain Sciences* 42 (2019): https://doi.org/10.1017/S0140525X19001006.

68 Ivan Colagè – Francesco D'Errico, "Culture: The Driving Force of Human Cognition", *Topics in Cognitive Science* 12 (2018) 654-672: https://doi.org/10.1111/tops.12372.

view of human evolution, which must be added to the previous, more traditional bottom-up view". In my view, this absence of genetic change is perfectly applicable to recent and minority abilities (e.g. writing, which these authors focus on) which are caused by individual reuse of older capacities. In some other cases, however, the influence of culture on human genome should not perhaps be despised. A fascinating datum is provided by Simon Neubauer *et al.*: Within the lineage of *H. sapiens* and in dates even later than 150.000 b. p. (that is, within the so-called "anatomically modern humans") there is a marked evolution in the shape of the cranium.[69] It is possible that such evolution reflects the influence of cultural and communicative innovations on genome. So far, what is clear is just that we will have to wait for future advances in human Paleogenomics and Genomics to be able to specify better what that evolution within *H. sapiens* was about, or, in other words, which "universal genetic inheritances" could arise within our species.

But let us return to the issue of this work. Here, we only need to defend that, originally, both inner speech and its role in complex self-control were adaptively advantageous novelties. And in order to defend that claim, it is not strictly necessary that those new "reuses" of social language are already set in the genome, but we can admit that, so far at least, it is only a reuse that universally appears in each individual brain. Certainly, unlike writing, those abilities are universal within our species, and they are also much older than writing. But, despite that, it might be enough for the genome to maintain the universal human ability to learn the language—this requirement might be enough— to get any human individual during normal ontogenesis to reach the neural reuse which mature inner speech entails. Thus, one way or another, that inner speech and its role in complex self-control are an evolved, originally advantageous resource.

---

[69] Cf. Simon NEUBAUER – Jean-Jacques HUBLIN – Philipp GUNZ, "The Evolution of Modern Human Brain Shape", *Science Advances* 4 (2018): https://doi.org/10.1126/sciadv.aao5961.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

6. THE TWO TYPES OF SELF-CONTROL

Since inner speech is useful for any type of "complex self-control" (or "willpower"), we must compare the self-control exercised in favour of one's own goals and the self-control involved in the base for the most demanding moral capacity. The main difference is that only the latter type *necessarily* involves the grasping of somebody else's mental contents. It is the truth of foreign mental needs that motivates there the subject and leads him to face his own goals.

Another difference is that for the latter type, the minimum essential requirements —the easiest grasping of "foreign" mental needs and the mature inner speech— are nowadays universal in the strongest sense, or, in other words, they are similar in all human adults who are normally constituted. However, complex 'selfish self-control' involves some requirements (ability to give sensory vividness to the particular evocations which are of interest,[70] agility of working-memory, and —sometimes— a highly recursive theory-of-mind of the sort "He thinks that I think that she…") which present a great variability of degree amongst human individuals.[71]

But there is a more important question in all this issue. We saw said that, if the subject has derived an action just from goals of his, then such action cannot be linked to the most demanding moral capacity but to selfish self-control. However, throughout history, many people have believed —and believe— that good actions will be rewarded in the afterlife. Is

[70] The role of vivid evocations in reducing delay discounting (that is, in increasing the choice of "the later but better reward") continues today to be corroborated by new studies: Leonard Epstein *et al.*, "A Story to Tell: The role of narratives in reducing delay discounting for people who strongly discount the future", *Memory* 29 (2021) 708-718: https://doi.org/10.1080/09658211.2021.1936560. But such influence of narratives can also be applied to the really moral self-control. The idea of that application was pointed out by an anonymous reviewer.

[71] We might add that one's ontogenesis must have promoted "internal *vs*. external locus of control". Cf. Julian B. Rotter, "Generalized Expectancies for Internal Versus External Control of Reinforcement", *Psychological Monographs: General and Applied* 80 (1966) 1-28: https://psycnet.apa.org/doi/10.1037/h0092976.

the most demanding moral capacity necessarily absent in all those individuals?

At first sight it seems that it has to be absent. Complex self-control —we have admitted— takes long term into account. How can we then expel the search for posthumous rewards from selfish self-control? Do not these rewards continue to be bribes? But let us not rush to judgement. It is convenient to examine this matter further.

First of all, I accept that "the fear of punishments for having broken the group's rules" is adaptive for the individual, since the expulsion from the group could be lethal. And what is more, that adaptive advantage might become more intense when the fear takes the subject to stop his punishable behaviour even when the subject believes that no one is paying attention to him (Baumard *et al.*, which call this latter type of fear "genuine moral sense", highlight that the error of mistakenly assuming that no one is paying attention to a blatantly selfish action may compromise an agent's reputation[72]). This particular fear certainly contributes to stop crime and mischief that are easily tagged. However, it seems less plausible that such fear, which is a "distillation" of social control, can promote what I have called the most demanding moral capacity.

Going on to the rewards, we have to consider that in Christian religion posthumous rewards are said to be different from the most typically coveted goods — to be almost unknown (1 Cor 2,9-10).[73] And it is regarding such special rewards that we have to ask ourselves: How could the belief in them get the subject to try to adopt a decision which is annoying and

---

[72] Nicolas BAUMARD – Jean-Baptiste ANDRÉ – Dan SPERBER, "A Mutualistic Approach to Morality. The evolution of fairness by partner choice", *Behavioral and Brain Sciences* 36 (2013) 59-122: https://doi.org/10.1017/S0140525X11002202. These authors propose that "competition between individuals became very intense to be chosen as a partner in cooperative ventures" (p. 65).

[73] At the beginning, the afterlife did not have the role of rewarding the actions that the individual would have performed throughout his life, but it was considered the mere consequence of some degree of continued existence. But, afterwards, new ideas paved the way —that of heavenly rewards to good actions and, later, that of different, almost unknown character of such rewards.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

contrary to the goals which are active in him at the moment? The possibility that such belief can achieve such a thing seems very strange when it is considered that those unearthly rewards would be completely disconnected from the biological forces which are active in us in the roar of each of those situations. And that sense of astonishment becomes reinforced when we observe that the long-term goals which support selfish self-control are, however, always connected in one way or another with biological forces similar to those which in the short term are opposed to that self-control: For instance, it is from our very concrete hopes for prestige and power (or, in children, for positive evaluation by parents[74]), where selfish self-control draws strength when it demands hard efforts from us.

However, is it correct to assume that such disconnection —such separation between the almost unknown rewards and what is experienced in those situations— necessarily happens? I would answer that the reality of the disconnection may depend on what the subject is doing. Let us suppose that the subject is trying to place a "foreign" interiority at the same level as his own or, in other words, to make that similarity —that truth that he perceives— to have a real influence on his own behaviour. In that case (in analogy with that goal that he is pursuing, but also as a contrast with the impossibility of his task[75]) the —received— idea of a completely wise and fair mind will finally be activated in him, a mind which understands thoroughly and generously the interiority of all individuals — and, therefore, also understands the subject as he, the subject, has always longed to be understood.

Let us move on to see the implications for our matter. It would be through the process of trying to make that type of decisions, or, more concretely, throughout the ideas which are activated there —firstly, that of the supra-human mind,

---

74 Laureano Castro – Miguel Ángel Castro-Nogueira – Morris Villarroel – Miguel Ángel Toro, "The Role of Assessor Teaching in Human Culture", *Biological Theory* 14 (2019) 112-121: https://dx.doi.org/10.1007/s13752-018-00314-2.

75 Analogy and contrast between God and man: "Similarity within a greater dissimilarity" (a description applicable to any human capacity — e.g. "mental time travel" mentioned in note 20).

and then, already unstoppable, that of the approving look of that mind at the current effort of the subject—, how those beliefs could finally connect with the subject's experience at that given moment. In short, at the beginning, we were taking for granted that the belief in rewards of a heavenly, almost unknown type would be the cause for the most demanding moral decisions, but now instead we suggest that the ability of that belief to influence such decisions would be caused by the attempt to make them. It is only at those moments that the belief in rewards of a special, heavenly type becomes connected with the situation and can thus truly leave a mark and have an influence.

7. TAKING A CLOSER LOOK AT SOME CONSEQUENCES OF THE ADVANCED MODE OF ToM

In the previous section, we have not only answered an objection, but also found a new issue. The completely wise and fair mind understands the subject with the depth with which he, the subject, has always longed to be understood. That longing or yearning is the late culmination of self-conscious emotions.

It is because of this yearning that the observation of a (small or large) group where all its members help and love each other —that mere observation— works as a true causal boost for the observer to want to be a part of such type of group. In addition, there, the observer does not feel that the observed actions of costly help would be annoying for those who carry them out. There, to the eyes of the observer, the loving support and gratitude that each of those individuals gets from the others seems to extend a light full of attractiveness on the costly actions, and such light does not depend on any long-term goals. In short, that model of continuous cohabitation will surely dazzle any observer.

However, when the observer tries to be a part of such a type of group, he will immediately notice that he has not got as much support as he at each moment expected, and he will find out that the longed-for cohabitation imposes on him demands

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

which are annoying and almost unbearable.[76] The causes for that disappointment are clear. On the one hand, we all long to feel fully understood and loved by an ideal ability, but, on the other hand, our ability for each of us to understand and love others is poor and typically very poorly exercised.

Nonetheless, in spite of that disappointment, the longing remains in all of us, and some confidence in the possibility of its fulfilment can help us along the way of real cohabitation. Regarding the need for such help, note that "anyone who wishes to give love must also receive love".[77] And note also that real cohabitation is a very difficult way. There, the subject must sustain the others, instead of being, like in his previous imaginations, sustained by the others: he must make the others' moral decisions easier, instead of receiving from the others such facilitation (cf. Lk 10,29.36). "In these situations, we are called to be living sources of water from which others can drink. And at times, this becomes a heavy cross".[78]

But taking into account what I have just said ("some confidence that our longing could be fulfilled favours moral decisions in us"), does that not mean that I am, against my previous claims, including the most demanding moral capacity within "selfish self-control"? I would reply that I do not disown my posture. That longing for achieving ideal cohabitation is not selfish but precisely the opposite. That longing tends, in its last resort, for each mind to take in generously the foreign mind. It could be said that in ideal cohabitation each subject, after the

---

[76] This non-ideal behaviour on the part of the group causes in the subject a weakness of the moral will, even if no actual fault can be imputed to that subject. See Ratzinger about original sin: "When the network of human relationships is damaged from the very beginning, then every human being enters into a world that is marked by relational damage. At the very moment that a person begins human existence, which is a good, he or she is confronted by a sin-damaged world". Joseph Ratzinger, *In the Beginning: A Catholic Understanding of the Story of Creation and the Fall*, Grand Rapids, MI – Cambridge, Eerdmans, 1986, 73.

[77] Pope Benedikt, *Deus caritas est*, 7.

[78] Cf. Pope Francis, *Evangelii Gaudium*, 86. In other words, we are called to "fill up in our own flesh what is still lacking in regard to Christ's afflictions" (Col 1,24). It is in this way that the joyful healing of hurt relationships could (at least partially: Mk 10,30) reach the interhuman level too.

spirit of love has done its job, mirrors the foreign mind with his own mind, but not because previous differences are faded, but because each has managed to become the total of both: There, the "as yourself" is a description, not a precept.[79]

But, once I have defended the unselfish character of such longing, let us put our feet back on the ground and focus on our real cohabitation. The decisions of true moral freedom are connected with the annoying little details of everyday cohabitation.[80] In addition, and returning to the general proposal, that longing, as well as true moral freedom, is a consequence of what I have called the advanced mode of ToM.

8. A FINAL SUMMARY

Sections 2, 3 and 4 propose an evolutionary base for the most demanding moral capacity. This base, while it is not an adaptive advantage but quite the opposite, arises from the convergence between two abilities which in their respective origins were adaptively very advantageous. Originally, the ability of grasping somebody else's mental contents (vs. merely activating vicarious expectations) was linked to self-conscious emotions and probably also to the human type of communicative reception. Likewise, in its beginning, the second ability —the speech directed to oneself— provided the strong self-control that became required when the gossip converged with self-conscious emotions.

But, as stressed in Introduction, this base is too poor and weak to sustain the exercise of the most demanding moral capacity. This is why, according to Christian religion, human beings require God's grace. Here we enter a completely different terrain, that of revelation and supernatural gifts. However, supernatural gifts fit well with the natural processes linked to the base.[81] Sections 6 and 7 focus on that fit.

---

79 Could this be in the image of the communion of Trinitarian life?

80 Except a special type of moral decisions: see above, note 56.

81 It goes without saying that such a fit, like any other fact in the universe, can be regarded as either favourable or contrary to faith: To choose between the two possibilities is one's own decision.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056

When does the revelation begin to have a weight on our situation? Section 6 proposes that it is when we experience the difficulty or, rather, impossibility of really loving others. More concretely, it is only then that the idea of a completely wise and fair mind that loves us all and the idea of special supernatural rewards can motivate us effectively.

We can also look for that fit from another angle. This is what section 7 tries to do. The human being longs for an ideal group in which he receives all the understanding and affection and gratitude he expects and where, as a result, he has no difficulty in loving. This longing arises as a derivation and deepening of the self-conscious emotions and thus belongs to the natural base. Of course, all this immediately crashes into reality. However, it can also be said here that the revelation —the supernatural hope that we will achieve the longed-for ideal group— fits well with the base that the complex detours of evolution have provided us with.

ISSN: 1131-7027 ISSN-e: 2660-7743
https://doi.org/10.46543/ISID.2231.1056 ISIDORIANUM 31/2 (2022) 91-126