Sánchez, Á. (2018). Physics of human cooperation: experimental evidence and theoretical models. *Journal of Statistical Mechanics: Theory and Experiment*, 2018.

DOI: https://doi.org/10.1088/1742-5468/aaa388

Physics of human cooperation: Experimental evidence and theoretical models

Angel Sánchez[1,2,3,4]

[1] Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain

[2] Unidad Mixta Interdisciplinar de Comportamiento y Complejidad Social (UMICCS) UC3M-UV-UZ, 28911 Leganés, Madrid, Spain

[3] UC3M-BS Institute of Financial Big Data (IFiBiD), Universidad Carlos III de Madrid, Calle Madrid 135, 28903 Getafe, Spain

[4] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, 50018 Zaragoza, Spain

E-mail: `anxo@math.uc3m.es`

November 2017

Abstract. In recent years, many physicists have used evolutionary game theory combined with a complex systems perspective in an attempt to understand social phenomena and challenges. Prominent among such phenomena is the issue of the emergence and sustainability of cooperation in a networked world of selfish or self-focused individuals. The vast majority of research done by physicists on these questions is theoretical, and is almost always posed in terms of agent-based models. Unfortunately, more often than not such models ignore a number of facts that are well established experimentally, and are thus rendered irrelevant to actual social applications. I here summarize some of the facts that any realistic model should incorporate and take into account, discuss important aspects underlying the relation between theory and experiments, and discuss future directions for research based on the available experimental knowledge.

Keywords: Complex systems, evolutionary game theory, cooperation, behavioral experiments.

1. Introduction: Physics and the social sciences

While the idea of applying physics methods to social phenomena goes back some two hundred years (see [1] and references therein), the coming of age of statistical physics as the proper description of complex systems has stimulated a rapid development of the field in the last two decades. Physics journals are now populated by thousands of models that aim to describe this, that and the other in socio-economic phenomenology. Just as an example, the review *Statistical physics of social dynamics* [2], published in 2009, contains hundreds of references to papers in this area, and many more have appeared in subsequent years. Unfortunately, in contrast to biology, where physicists work very often hand in hand with biologists and, in particular, with experimental biologists, those papers are quite far from applicable in the social sciences. Quoting Castellano *et al.*:

> The introduction of a profusion of theoretical models has been mainly justified by vague plausibility arguments, with no direct connection to measurable facts. Very little attention has been paid to a stringent quantitative validation of models and theoretical results. The contribution of physicists in establishing social dynamics as a sound discipline grounded on empirical evidence has been so far insufficient.

Subsequent work does not seem to have taken care of this problem. The amount of theoretical knowledge accumulated by physicists is already enough to be collected in textbooks such as [3], but it still criticized from the social sciences camp for its obliviousness to their contributions [4]. In this mini-review I point out to one of the reasons for the lack of contact between social research done by physicist and real-life issues: The fact that empirical and experimental knowledge is rarely incorporated into models. However, a remark is in order before going into this discussion. There are two *bona fide* motives to study a model of cooperation from the physics viewpoint: exploring and explaining real observations of cooperation (the one that this paper deals with) and studying novel statistical mechanics in the sense of e.g., finding new universality classes with consequences at the collective behavior level. "Classical" or "canonical" models, such as the voter [5], Schelling [6] or Axelrod [7] models are good examples of papers on physically interesting research that hints on actual social phenomena (see also [2]). Thus, a model that has no direct connection with social phenomena may still be interesting but, clearly, only a minority of the "profusion of theoretical models" mentioned above falls in this class. For the rest, lacking theoretical relevance, only their relation to actually observed features of cooperation may make them appealing to the social science community at large.

2. Physics of cooperation

As an appropriate example of the shortcomings of this research, I will focus below on the issue of the emergence and sustainability of cooperation. Why individuals cooperate with their conspecifics is a general puzzle of evolution, in so far as helping others diminishes the chances to transmit one's genes to the next generation. This is the case with all living beings, from virus and bacteria [8] to humans [9] through insects [10]. In our case, humans are considered to be hyper cooperative, i.e., we cooperate to a much larger extent than our relatives, the great apes [11], leading to our unusual cognition, technology and culture [12]. Therefore, it is no surprise that the questions as to how cooperation arises and survives has attracted the interest of researchers in many fields, from economics (e.g., [13] to biology [14], and of course, of many physicists (who sometimes make the mistake of neglecting contributions from other fields [15]). Thus, a very recent review paper [16] on the statistical physics of human cooperation features 295 references, in what is very likely a lower bound on the number of papers dealing with this problem.

Much of the research on the evolution of cooperation relies on evolutionary game theory [9, 17, 18] and, in particular, in the Prisoner's Dilemma and related games [19]. In this stylization of the difficulties arising in cooperation, two people have to decide whether they want to cooperate with the other or not. Mutual cooperation provides both of them with a benefit, but it is more convenient not to cooperate, so one benefits from the other's cooperation at no personal cost. Thus, the best option is skip cooperation, but when both agents think in this manner and choose this option no benefit is produced and they turn out worse than if they had both cooperated, hence the dilemma. It is clear that this is an abstract but faithful representation of the puzzle of cooperation and, as such, has been studied at great length. To add evolution to this game-theoretical setup, a population of agents is considered: these agents are endowed with a strategy, i.e., a recipe to decide whether to cooperate or to defect, and interact repeatedly among themselves. Those who benefited more from the interaction are more likely to reproduce and be represented in the next generation, the specific rule governing reproduction being one of the model ingredients.

It is probably fair to say that the pioneering paper of this research was the one on evolutionary games and spatial chaos by Nowak and Sigmund [20]. This was a cellular automata model [21] intended to explore nontrivial dynamics of spatially extended systems in physics and biology but, as a side result, showed that cooperation could survive when agents were located on the nodes of a square lattice and interacted only with their neighbors. This result spawned many other works dealing with evolutionary game theory on graphs [22] and also led to the general belief that cooperation was promoted by the existence of a (lattice or network) structure on the population. However, it was later shown [18] that this was by no means a general result, and the survival of cooperation depended on the model details: The composition of the population in terms of the strategies present, the initial conditions, the reproduction rules, the population structure and its details (particularly when considering networks), etc. In such an unclear scenario, more and more models were proposed "justified by vague plausibility arguments", and almost no research tried to make a connection with reality. A reality which, as always in physics, eventually comes down to empirically established facts through experiments.

3. Cooperation on structured populations: empirical facts

Indeed, one particularly relevant source of information for research on cooperation (and on any social interaction, for that matter) is behavioral economics [23]. This discipline, that became accepted as mainstream economics with the award of the Nobel

Memorial Prize in Economics to Kahneman and Smith in 2002, is the equivalent to physics experiments as far as validation of game-theoretical models is concerned. What researchers typically do is to recruit a number of subjects (more often than not, university undergraduate students) that are asked to play the game of interest (in an anonymous setting to avoid confounding influences) and monetarily rewarded according to everybody's decisions during the experiment. Arguably, this is a highly idealized version of real life interactions but, on the other hand, the same could be said of a majority of physics experiments that are carried out in laboratory setups. In fact, it is sometimes possible to take these experiments to more realistic contexts [24], although the logistic difficulties are considerable and sometimes unsurmountable.

Within behavioral economics, the issue of cooperation on structured populations has been considered both by economists [25, 26] and by others [27–31]. A disclaimer is in order here as to the extent of the literature on this field: Research on cooperation is practically ubiquitous and appears in fields ranging from psychology to computer science, and it is a herculean task to follow all these many venues, so this list should probably be complemented with some other papers. Even with this caveat in mind, from the experiments the following conclusions regarding empirical facts can be drawn:

Lattices or networks do not support cooperation. With the exception of [31], most papers [25–30] find that cooperation starts at a high level, with more than half the agents in the population choosing to cooperate, and proceeds from there to a more or less slow decline towards a small (around 20% or less) fraction of cooperating agents. This is in fact the general behavior observed in the general class of games known as "public goods games" (see [32] for a review). Regarding the experiments in [31], they address the claim that when the ratio between the benefit of cooperation and the cost of cooperating exceeds the number of neighbors, cooperation should be supported. This typically means that the cost of cooperation is very small and the other quoted papers are in the opposite regime, which would explain the negative result. On the other hand, when cooperation is cheap, it is clear that it will also be easier to maintain. However, the experiment in [31] is problematic on its own, because different replications (in the lab and using Amazon Mechanical Turk workers) of the experiment reported in the paper yield widely different results (80% vs 40% cooperation), and because the fraction of cooperation is not monotonic on the control parameter, contrary to the predictions. These and other issues, such as the short duration of the experiment compared to others, which would warrant a much longer discussion, cast some doubt on this result, although of course it has to be kept in mind as we will come back to later.

People are moody conditional cooperators. Conditional cooperation refers to a strategy that is more likely to cooperate when more neighbors around cooperated with oneself, and was discovered in [33] in the context of unstructured group interactions. For structured populations, it was found in [27], where it was also discovered that the previous action of the subject also conditioned her choices, being more likely to cooperate after having cooperated and vice versa. This was further confirmed in the experiments [29, 30, 34, 35]. Interestingly, when one assumes that this is the way people choose their actions when playing a Prisoner's Dilemma on a structured population, it has been theoretically shown that the behavior of cooperation does not depend on the structure considered [36], which would indicate that lattices or networks are no different from an unstructured population in agreement with the experimental observations [29].

People do not take into account the earnings of their neighbors. This is a result arising from the meta-analysis of three different experiments in [30], and it implies that the experimental subjects take their decisions based on actions only (according to a moody conditionally cooperative strategy as described in the previous paragraph) and, perhaps, on their own benefit of the interaction. This is a very important finding: Indeed, many evolutionary models assume that agents that obtain more benefit from the interaction are more represented in the next generation. This might very well be the case in other living beings (such as bacteria, for instance [8]) but our result for humans goes against the applicability of this kind of updating to describe human social phenomena. This is also in agreement with the intuition that when we cooperate with others, we know what others did to us (cooperate or not) but we are generally unaware of the specific benefit they reaped from the interaction. In this respect, it is important to note that people still may be mimicking others, i.e., copying their actions, but they choose which action to copy based on criteria that do not involve the payoffs others obtain with their choices. Thus, forms of conditional cooperation may arise from copying the action of the majority, or from choosing to cooperate with probability proportional to the number of cooperating neighbors, but in none of these the payoff of the others is considered.

Cooperation can be sustained in dynamic networks. When, instead of playing with fixed neighbors on some structure, subjects are allowed to modify their network at will (unrestrictedly or with certain constraints), cooperation is promoted, and stable fractions of cooperators can be 60% or above [37–44]. In essence, this is nothing but assortment, i.e., the facts that cooperators get together to benefit each other, leaving out those that defected. This was proposed as the true and only ultimate reason supporting cooperation in [45], which would translate into different proximate mechanisms of social relationship. In the case of cooperation on structured populations, it appears that information on the others' behavior (but, again, not on their benefits) or, in other words, reputation, is the mediator for the choices of whom to connect to [40], although it has been found in more elaborate cooperation games that reputation does not always help [46]. Interestingly, the mechanism underlying the stability of cooperation in dynamic networks may be more general, as it can be interpreted as punishment: When my neighbor defects and benefits from my cooperative action, I punish her by cutting the corresponding link and abstaining from interacting. Such behavior is indeed related to the punishment that has been observed to sustain and even increase cooperation in public goods games; in this case, players are allowed to pay to have money taken from others [47] which is a similar manner to inflict an economic damage to one's counterparts. Conditional cooperation might be also interpreted as a manner of punishment in this context. For dynamic networks, thinking in terms of punishment would also explain the relevance of the past history or reputation in order to decide on which links to cut. Interestingly, this interpretation agrees with the theoretical results in [48], where the importance of credible, player-specific punishment to support collaboration on networks is highlighted.

Having summarized the facts we have learned about cooperation on structured populations, a word is in order to discuss how they apply to real life and to models. My contention in this paper is that models are accepted (always provisionally) or rejected based on the results of experiments. To that end, as in physics, experiments are designed and carried out trying to falsify the predictions of a model and, as such, they should be as close to it as possible. One obvious physical example is the research done with the LHC at CERN, far from any real life situation, but in spite of that it is the ultimate test of any high-energy physics theory. Then, it may turn out that the model predictions are true as in the claim above, but it also occurs (and very often) that the model predictions are wrong: For instance, there are tenths of theoretical papers claiming that scale-free networks should give rise to large values of cooperation, but the available experimental evidence, albeit scarce, does not support this claim. This is so because, contrary to the comment above, people do not necessarily behave in a self-centered or selfish manner, and therefore the outcome of an experiment is more often than not surprising even to the experimenter.

In fact, once a the mechanism considered in a model has been experimentally demonstrated, the next question is then whether or not it applies in real life. Here it is important to consider the differences between models and real human behavior. One such difference is that many models consider agents that choose their actions in the game without taking into account that it will be repeated. Indeed, a rational player interested only in maximizing her own monetary payoff should play according to the game theoretical concept of Nash equilibrium [49] when the game is played only once: In one-shot Prisoner's Dilemmas, for instance, this corresponds to defection. However, when the game is repeated an infinite number of times or players do not know the exact number of rounds to be played many equilibria (including some cooperative ones) are possible [17, 50]. A related issue is the fact that people recognized that they are playing with the same counterparts and that they may punish them as discussed above, even at a personal cost, to ensure cooperation in the future, something that relies on a theory of mind [51] that agents in most models do not have. Experiments are, therefore, key to understand this type of problems as they are done with real people that may not behave like the agents featured in the models, but their results have to be interpreted in the light of such dissimilarities.

4. Concluding remarks

In this paper we have seen that there are many relevant experimental results on cooperation on structured populations published in widely read journals while, unfortunately, many models are introduced in the literature without taking into account the facts established above. As physicists, we learned that knowledge only advances when theory and experiment go hand in hand, feeding back on each other. In the case of the physics of human cooperation, we are at the stage that only models which are compatible with the known facts should be studied and published, and this is certainly not the case. To be sure, I stress that I am discussing human cooperation here; there are many other applications, from bacteria to bots, where models that do not abide by the empirical facts summarized above may be correct descriptions of the real world.

Having said this, it appears that there are two main directions for research starting from the empirical facts. An important one is to clarify the possible dependence of the emergence of cooperation on the parameters of the game, as claimed in [31]. In this respect, further experimental research is needed. One important concern at this point is the use of Amazon Mechanical Turk workers as experimental subjects: the fact that a pool of some 30 000 turkers takes part in more than 1000 experiments per year [52] and that their behavior depends on many external factors such as the time of the day [53] is certainly problematic. As in any other experimental field in the social sciences, replications are badly needed [54, 55] (sadly, reproducibility is also an issue in better established fields, such as, e.g., cancer research [56] where initiatives to solve this problem are underway [57]); similar efforts would be important for a better understanding of cooperation. On the other hand, international collaborations that allow to compare the behavior of people from largely different backgrounds are also indispensable to assess the generality of the results. The second direction in which research would contribute to advance the field is theoretical: Models incorporating the empirical facts and otherwise being specific on other aspects of behavior (such as, for instance, dependence on social characteristics of the participants [58]) should be developed to gain further insight on the interplay between social structure and cooperation. Their predictions could then be tested on new experiments specifically designed to that end, thus establishing further empirical facts. A very interesting issue is that of system size: It is quite obvious that a system with a few people may behave very differently from a system with thousands of interacting agents. One, but not the only reason for that could be the amount of information one has to process in a very large system as compared to a small one. The very physical issues of scaling (such as the growth of the number of possible equilibria with the system size [48]) and phase transitions will surely play a role in such research. Progress in these and related directions can only be achieved by the judicious combination of theoretical and experimental efforts, which I hope will be stimulated by this mini-review.

Acknowledgments

I am most grateful to all my collaborators in the experiments I have participated in since my first one [27]. I have learned a lot from and with them, from the experimental design phase to the analysis of the resulting data. This work was partially supported by the EU through FET-Proactive Project DOLFINS (contract no. 640772) and FET-Open Project IBSEN (contract no. 662725), and by Ministerio de Economa y Competitividad (Spain)/FEDER (EU) grant VARIANCE (contract no. FIS2015-64349-P). Project IBSEN is in fact about large-scale experiments and provides a open-access platform for researchers to carry out them [59].

References

[1] Stauffer D 2012 *J. Stat. Phys* **151** 9–20
[2] Castellano C, Fortunato S and Loreto V 2009 *Rev. Mod. Phys.* **81** 591–646
[3] Sen P and Chakrabarti B K 2013 *Sociophysics: An Introduction* (Oxford: Oxford University Press)
[4] Troitsch K G 2014 *J. Artif. Soc. Social Simulation* **17** issue 1
[5] Clifford P and Sudbury A 1973 *Biometrika* **60**, 581–588
[6] Schelling T C 1971 *J. Math. Soc.* **1** 143-186
[7] Axelrod R 1997 *J. Conflict Resol.* **41**, 203–226
[8] Wingreen N S and Levin S A 2006 *PLOS Biol.* **4** e299
[9] Axelrod R 1984 *The evolution of cooperation* (New York: Basic Books)
[10] Wilson E O 2000 *Sociobiology: The New Synthesis, Twenty-Fifth Anniversary Edition* (Cambridge: Belknap Press)
[11] Cronin K A 2012 *Anim. Behav.* **84** 1085–93
[12] Tomasello M and Vaish A 2013 *Ann. Rev. Psychol.* **64** 231–55
[13] Bowles S and Gintis H 2011 *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton: Princeton University Press)
[14] Wilson E O 2000 *Sociobology* (25th Anniversary Edition, Cambridge: Harvard University Press)
[15] Durlauf S N 1999 *Proc. Natl. Acad. Sci. USA* **96**, 10582–10584
[16] Perc M, Jordan J J, Rand D G, Wang Z, Boccaletti S and Szolnoki A 2017 *Phys. Rep.* **687** 1–51
[17] Gintis H 2009 *Game Theory Evolving, 2nd Edition* (Princeton: Princeton University Press)
[18] Roca C P, Cuesta J A and Sánchez A 2009 *Phys. Life Rev.* **6** 208–49
[19] Rapoport A and Chammah A M 1965 *Prisoner's Dilemma* (Ann Arbor: University of Michigan Press)
[20] Nowak M A and Sigmund K 1992 *Nature* **355** 250–53
[21] Wolfram S 1983 *Rev. Mod. Phys.* **55** 601–44
[22] Szabó G and Fáth G 2007 *Phys. Rep.* **446** 97–218
[23] Camerer C F 2003 *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton: Princeton University Press)
[24] Sagarra O, Gutiérrez-Roig M, Bonhoure I, Perelló J 2016 *Front. Phys* **3** 93
[25] Cassar A 2007 *Games Econ. Behav.* **58** 209–30
[26] Kirchkamp O and Nagel R 2007 *Games Econ. Behav.* **58** 269–92
[27] Grujić J, Fosco C, Araujo L, Cuesta J A and Sánchez A 2010 *PLOS ONE* **5** e13749
[28] Traulsen A, Semmann D, Sommerfeld R D, Krambeck H J and Milinski M 2010 *Proc. Natl. Acad. Sci. USA* **107** 2962–6
[29] Gracia-Lázaro C, Ferrer A, Ruiz G, Tarancón A, Cuesta J A, Sánchez A and Moreno Y 2012 *Proc. Natl. Acad. Sci. USA* **109** 12922–6

[30] Grujić J, Gracia-Lázaro C, Milinski M, Semmann D, Traulsen A, Cuesta J A, Moreno Y and Sánchez A 2014 *Sci. Rep.* **4** 4615
[31] Rand D G, Nowak M A, Fowler J H, Christakis N A 2014 *Proc. Natl. Acad. Sci. USA.* **111** 17093–8
[32] Ledyard, J O 1995 In *The Handbook of Experimental Economics*, edited by J H Kagel and A E Roth (Princeton: Princeton University Press)
[33] Fischbacher, U, Gaechter S and Fehr E 2001 *Econ. Lett.* **71** 397–404
[34] Gutiérrez-Roig M, Gracia-Lzaro C, Perell J, Moreno Y and Snchez, A 2014 *Nat. Communs.* **5**, 4362
[35] Horita Y, Takezawa M, Inukai K, Kita T and Masuda N 2017 *Sci. Rep.* **7** 39275
[36] Gracia-Lázaro C, Cuesta J A, Sánchez A and Moreno Y 2012 *Sci. Rep.* **2** 325
[37] Ule, A 2008 *Partner Choice and Cooperation in Networks: Theory and Experimental Evidence* (Berlin: Springer)
[38] Rand D G, Arbesman S and Christakis N A 2011 *Proc. Natl. Acad. Sci. USA* **108** 19193–8
[39] Wang J, Suri S and Watts D J 2012 *Proc. Natl. Acad. Sci. USA* **109** 14363–8
[40] Cuesta J A, Gracia-Lázaro C, Ferrer A, Moreno Y and Sánchez A 2015 *Sci. Rep.* **5** 7843
[41] Gallo E and Yan C 2015 *Proc. Natl. Acad. Sci. USA* **112** 3647–52
[42] Antonioni A, Tomassini M and Sánchez A 2015 *Sci. Rep.* **5** 10282
[43] Riedl A, Rohde I M T and Strobel M 2015 *Rev. Econ. Studies* **83** 737–67
[44] Antonioni A, Sánchez A and Tomassini M, 2015 *Sci. Rep.* **6** 27160
[45] Fletcher J A and Doebeli M 2009 *Proc. R. Soc. B* **276** 13-19
[46] Corten R, Rosenkranz S, Buskens V and Cook K S 2016 *PLOS ONE* **7** e0155703
[47] Fehr E and Gächter S 2002 *Nature* **415** 137-140
[48] Dall'Asta L, Marsili M and Pin P 2012 *Proc. Natl. Acad. Sci. USA* **109**, 4395–4400
[49] Nash J F 1950 *Proc. Natl. Acad. Sci. USA* **36**, 48–49
[50] Fudenberg D and Maskin E 1986 *Econometrica* **54** 533–554
[51] Premack D and Woodruff G 1978 *Behav. Brain Sci.* **4** 515–526
[52] Bohannon J 2016 *Science* **352** 1263–4
[53] Arechar A A, Kraft-Todd G and Rand D G 2017 *J. Econ. Sci. Assoc.* **3** 1–11
[54] Aarts A A *et al* (The Open Science Collaboration) 2015 *Science* **349** 943-951
[55] Camerer C F, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan T, Heikensten E, Holzmeister F, Imai T, Isaksson S, Nave G, Pfeiffer T, Razen M and Wu H 2016 *Science* **351** 1433–6
[56] Begley C G and Ellis L M 2012 *Nature* **483** 531–533
[57] Baker M and Dolgin E 2017 *Nature* **541** 269–270
[58] Cronin K A and Sánchez A 2012 *Adv. Complex Sys.* **15** Supp no 1 1250066
[59] http://www.ibsen-h2020.eu