Shaun P. Hargreaves Heap*
Varieties of Tribalism in the Laboratory

Abstract: This paper uses evidence from laboratory experiments to identify a variety of tribalisms. This is important because some tribalisms encourage zerosum thinking and others do not; and some are not developed by Buchanan. This, in turn, supplies new insights into Buchanan's project of identifying the kinds of environment that encourage his sense of moral progress. In particular, current levels of inequality become a significant barrier to moral progress not only because they create an economic form of tribalist zero-sum thinking but they also undermine the scope for positive-sum

Keywords: tribalism, in-group bias, zero-sum thinking, positive and negative discrimination

1 Introduction

Allen Buchanan argues persuasively in *Our Moral Fate* (Buchanan 2020) that an evolved moral sense is not destined to be tribal. Our moral sense responds to our social environment: "some social environments stimulate the tribalistic potential of our moral nature; others stimulate the potential for inclusion" (xi). With this established, he sets about the main purpose of the book: to understand these connections so as to take control of our moral fates:

> I want to understand how moral progress comes about—especially progress toward greater inclusion and away from tribalism; and I also want to understand the reverse process, how people whose moralities are inclusive can regress to tribalism. But that is not my ultimate purpose: I want to understand moral progress and regression because I'm convinced that doing so will begin to provide the information we need to shape our social environment so that it fosters progress rather than limits or erodes it. (Buchanan 2020, xiv-xv).

In this paper, and in support of this general project, I examine what laboratory experiments tell us about tribalism. For this purpose, I use the definition of tribalism that he gives in the Introduction to *Our Moral Fate*.

---

*Corresponding author: Shaun P. Hargreaves Heap,** Department of Political Economy, King's College, London, UK, e-mail: s.hargreavesheap@kcl.ac.uk

> Notice that tribalism here doesn't just mean that it is our nature to distinguish between Us and Them; it means that we are biologically hardwired or programmed to relegate Them to a markedly inferior moral status or to exclude them from the circle of moral regard altogether. (2)

Thus, tribalism has two elements: a distinction between 'them' and 'us' and the treatment of 'them' less well than own-group members. In the experimental literature, this is known as the in-group bias.

I shall argue that the laboratory evidence on the in-group biases reveals a variety of tribalisms in the sense of different treatment of own and other group members; and this evidence points to possible changes in the 'social environment', conducive to Buchanan's sense of moral progress, that are not developed/highlighted in his book. This is the contribution of the paper: it elaborates on how Buchanan's moral progress can or might occur through an expanded understanding of tribalism.

I begin with a matter of translation. The laboratory evidence on the in-group bias concerns differences in the treatment of own as compared with other group members. It does not identify whether such differences arise from a perceived difference in the moral status of own and other group members. They are simply behavioural differences. This raises a question, then, as to how these laboratory insights might apply to Buchanan because his argument (see above) is ostensibly specifically concerned with tribalism with respect to moral status. The bridge that I use to make the connection turns on the key role that zero-sum thinking plays in Buchanan's articulation of what tribalism means and why it is worrying.

> The tribalistic mentality sees things in black and white, good and evil—as a no-holdsbarred, zero-sum conflict between Us and Them, for the highest stakes. Tribalism transforms disagreement into mutual hatred, mild condescension into utter contempt. (vii)

The zero-sum thinking aspect of tribalism is also centrally what concerns Buchanan when he discusses intrasocietal tribalism in chapter 8. Within a society like the US, much of the tribalism that seems increasingly to characterize and potentially threaten democracy does not involve differences in civic status as such, although it may: rather it is the corrosive effect of tribalistic zero-sum thinking that needs to be combated. This is made explicit in chapter 9 when the argument moves forward to 'taking charge of our moral fates'.

> One often hears that to combat tribalism, we have to learn to listen to each other. That's good advice, but incomplete and taken by itself not very helpful. What's needed are institutions that pro-vide incentives for listening and for compromise, institutions that encourage people who disagree with each other not to operate in the zero-sum, winnertake-all mode. (246)

Thus I focus, but not exclusively, in what follows on when the in-group bias of differential treatment in the laboratory encourages zero-sum thinking and when it does not. The first aspect of the variety in tribalism that I develop in this paper directly relates to this distinction.

In particular, the in-group bias in the laboratory can arise either through negative discrimination against the out-group and/or positive discrimination in favour of own-group. The difference matters because the one encourages zero-sum and the other positive-sum thinking around the existence of group differences.

To see this, suppose the emergence of group identifications in a society creates tribalism in the sense that an in-group bias in behaviour is observed. The bias could arise in two distinct ways. First, relative to previous behaviours when there were no group distinctions, individuals now treat own-group members no differently than before but they treat out-group members worse. This is what I define as negative discrimination: indeed, one might say the out-group is now 'denigrated' relative to the natural reference point of the treatment they could have expected absent such group identifications. In this case, the emergence of group identifications encourages a form of zero-sum thinking in the sense that each member of a group does gain a new sense of its own identity but only because they treat others worse (i.e they impose a cost on the other group). Second, and in contrast, the in-group bias could arise relative to when there were no salient group distinctions because individuals now treat own group members better than before and the out-group members are treated no differently. This is what I call positive discrimination. In this case, the emergence of group identifications encourages a form of positive-sum thinking in the sense that each group member gains a new sense of their own identity and they treat each other better but there is no cost in terms of the worse treatment of out-group members. In short, the emergence of group identifications here produces a benefit for all.[1]

We know from the laboratory that some kinds of interaction that mix scope for cooperation with conflict yield negative discrimination and others positive discrimination. Only the former is the worrying kind of tribalist behaviour for the encouragement it gives to zero-sum thinking.

---

1 In the two game theoretic interactions that I discuss below. I shall make the connection between negative and positive discrimination and non-positive and positive-sum thinking formal in the following way. I shall show that in so far as the emergence of group identifications is accompanied by negative discrimination, the average value generated in the interaction falls and in so far as it is accompanied by positive discrimination the average value generated the interaction increases. In short, the emergence of groups shrinks the size of the pie in one case and expands it in the other. This is the material base for the claim that the one encourages zero-sum (indeed worse, negative-sum) thinking and the other positive sum thinking.

The second aspect of variety in tribalism revealed in experiments that I discuss relates to the economic based differentiations between 'them' and 'us'. Buchanan is largely quiet about differences in economic status and the 'rich and poor' kind of 'them and 'us'. His 'them' and 'us' distinctions largely turn on other, non-economic sources of identity: race, ethnicity, nationality, etc. He recognises that inequality may exacerbate these non-economic types of tribalist distinctions within a society because it increases the incentives of the rich to lever (non-economic, identity based) tribal differences to secure their own economic power, but he does not examine the particularities that arise from the economic form of tribalism: the tribalism of the 'rich' and the 'poor'. This is a significant omission because contemporary tribalism within liberal democratic countries is not simply marked by these identity differences. Identity politics in this sense may have become more prominent in these countries, but class politics has not disappeared. Again this is important to recognize because it provides new insights into how to combat contemporary zero-sum forms of tribalism.

Thus I develop two types of variety in tribalism. One is within the non-economic identities that are the central concern of Buchanan and I distinguish between whether they generate zero or positive sum thinking. The other is an economic based 'them' and 'us' that can arise between the 'rich' and the 'poor' and which is largely overlooked by Buchanan as a distinct form of tribalism.

In the *next section*, I develop the experimental evidence around the origin of the in-group bias. In *section 3*, I consider the economic or class form of tribalism in the lab. I conclude in *section 4* by developing an argument, on the basis of this evidence, with respect to the kind of institutions or social environments that would encourage Buchanan's type of moral progress.

2 The In-Group Bias

The in-group bias in pro-sociality is well known from experiments in psychology and economics. In this section, I focus on two of my own experiments. This is partly, of course, because they are mine but it is also because they are particularly well designed to reveal the different origins of the in-group bias and this is what is important for the argument in this paper.

In these experiments, groups are created artificially in the lab in the group identification treatments. The subjects are randomly assigned to either a red or a blue group and subjects know their own group identification and that of whoever they are interacting with. In the control, the subjects are not given these group affiliations, they just interact with fellow subjects, unadorned by red or blue identi-

fiers. In the experimental literature, this minimal group paradigm is often thought to provide a rather strong test of the in-group bias because the sources of group identification are minimal. If the bias arises here, how much more likely is it when attachments to the group have rich social and histories. It is important in the use of the lab evidence to be persuaded by such an argument that the results have 'external validity'. While I think this argument has merit. I have a different view on external validity that is worth mentioning in case this interpretation of the minimal paradigm results is unpersuasive.

My background assumption is that people typically tend to rely on a repertoire of decision rules when they confront new or difficult decision problems because working out what action might best satisfy one's preferences is often either impossible or too time consuming in these circumstances. Laboratory decision problems also frequently fall into this category of novel or difficult decision problems because they are often very abstract, have explicit money pay-offs, and are made anonymously. Thus, I believe it is likely that subjects rely in the lab on decision rules or habits of decision making that are triggered by features of the experiment. With this perspective, experiments are a vehicle for discovering the character of these rules or habits of decision making that are used on occasion both inside and outside the lab. This is why the lab insights can have external validity outside the lab. In this particular case, the extraneous feature of the experiment that is liable to trigger specific rules are the group identities and what is potentially revealed, to use a different vernacular, is the 'unconscious bias' of the subjects that is present both inside and outside the lab. It does not matter in what follows which interpretation of the experimental evidence is taken. What matters is that the results can be defended as having in one way or the other relevance for understanding behaviour outside the lab.

My first experiment involves a trust game (TG). There is a 1st and 2nd mover in this game. The 1st has an endowment and must decide how much (x) if any to 'give' to the 2nd mover. Whatever is 'given' is multiplied by 3 and the second mover must decide how much of this (3x) to 'return' to the first mover. Since a selfish rational choice 2nd mover will keep anything that is given, a selfish rational choice 1st mover will give nothing. This is the conventional game theoretic prediction. Thus in so far as the 1st mover does give something, they must believe that they can trust the 2nd mover not to act selfishly and return something (i.e. demonstrate trustworthiness). In this way 'giving' and 'return' rates of the 1st and 2nd movers respectively are treated as indexes of trust and trustworthiness in this game.

The second experiment is a two person public goods game (PG) where each person decides how much of an endowment to contribute to a public good (i.e. a continuous decision version of the classic prisoners' dilemma). The contributions are multiplied to produce a public good that is shared equally. The multiplicand is

less than 2 so that the selfish interest is to make a zero contribution (because half of something multiplied by less than 2 is less than the original contribution) and free ride on any contribution made by the other person. This is the conventional game theory prediction and as a result any contribution is typically regarded as evidence that players are not exclusively selfish: they must have some 'social' preference that is oriented to the interests of the other player and which motivates them to make a positive contribution. Or to phrase this slightly differently: the presence of these social preference sources of motivation transforms the interaction away from the classic prisoners dilemma interaction to, for example, something like an assurance game where it is rational for an individual to contribute to the public good.

Thus both experiments get subjects to play two-person games, both introduce group affiliation/identification in the same way and both allow easy identification of a subject's pro-sociality (i.e. their departure from rational choice selfish predictions). Furthermore both games are thought to capture the essence of many interactions in social life because they mix elements of conflict with that of cooperation.[2]

Table 1 gives the aggregate results for these two experiments. The first column gives the average Baseline giving rate in the trust game and the average contribution rate in the public goods game, where the Baseline has no group affiliations. Columns 2 and 3 refer to the own-group and other group rates (i.e. the rates when partnered with a member from own group or the other group) in the sessions with group affiliations (see Hargreaves Heap/Zizzo 2009 and Corr et al. 2015, respectively for the full details of each experiment).

	Baseline	To own-group	To other-group
Giving rate	0.56	0.56	0.35
Contribution rate	0.26	0.40	0.26

Tab. 1: TG 'giving' rates and PG 'contribution' rates

---

2 It is also, perhaps worth remarking that while these games are indeed thought to capture the essence of many social interactions now, they differ in this respect from the paradigmatic evolutionary interaction between groups that Buchanan discusses because these interactions typically only involve conflict.

The in-group bias is revealed in both experiments in the comparison between the 2nd and 3rd columns: the 'giving' rate is significantly higher to own than to other group members in the trust game and likewise the 'contribution' rate in the PG. However, the source of the bias is completely different. In the trust game, it comes from negative discrimination towards other-group members: the 'giving' rate to other-group members is significantly lower than the rate in the Baseline where there are no groups and the 'giving' rate to own-group members is the same as in the Baseline. Thus the introduction of group identification makes our subjects less trusting of other-group members than they would otherwise be. There is no gain to own-group trust. Own-group members get special trust recognition compared with other-group members but only because other-group members are trusted less.

To express this connection between negative discrimination and zero-sum thinking slightly differently: in so far as the emergence of group identifications occurs in a way that there are some 'between' as well as 'within' group interactions, then it follows that the aggregate level of trust will be lower with the emergence of groups and negative discrimination than before. Since the level of trust affects the productivity of the interaction (as this determines the amount that is multiplied by 3 in the trust game), this means the pay-offs generated in such interactions fall with the emergence of groups. This is the material base, as it were, behind the claim that the emergence of groups with this type of tribalism encourages zero-sum thinking. Group identification is associated with a reduction of the size of the pie.

In the PG, in contrast, the in-group bias comes from positive discrimination. The 'contribution' rate is the same when interacting with an out-group member as in the Baseline; and the in-group bias arises because the contribution when interacting with own-group members is significantly higher than in the Baseline. The introduction of group identities in the PG again gives special recognition to own-group members but it does not come with any cost for out-group members. They are not treated worse or denigrated through their group identification in the manner of the trust game.

Again, this point can be seen slightly differently by contrasting the size of the pie when there are no group identifications with when there are. It grows with the emergence of this kind of tribalism because in so far as there are some PG interactions within a group, their productivity increases and there is no change in the productivity of PG interactions between different group members. In short, this kind of tribalism brings a positive-sum benefit.

In some respects this experimental result may seem surprising because there is an emerging conventional wisdom in social science that heterogeneous societies (i.e. those where there are various distinct groups) suffer from less social capital in the form of trust and cooperation between people than do more homogenous ones

(where there are less or no such group differentiations).[3] Our experimental result supports the trust part of this conventional wisdom because in so far there are groups, then some interactions occur between groups where there is lower trust and this lowers the aggregate level of trust. But our experiment does not support the cooperation part of this conventional wisdom: we would expect more social capital of cooperation in diverse societies from our PG experiment.

The conventional wisdom is based on observational data, largely from cross section survey evidence that uses versions of the question 'Generally speaking, do you think you can trust people these days?'. Much less, if any, evidence really concerns the willingness to cooperate in public goods interactions. Indeed, Putnam (2007), who is one of the key figures in developing the analytic category of social capital through his famous book and article on 'Bowling Alone', finds that there is no evidence of this effect for cooperation, while it is there notably for trust in his cross section analysis of social capital in the US. Most of the authors in this emerging conventional wisdom have instead relied on theoretical arguments that trust underpins cooperation, with the result that clear evidence on trust and heterogeneity can be assumed to apply for cooperation and heterogeneity (e.g. see Collier 2018).

The theoretical argument employed by these authors has the ring of plausibility because it seems that to engage productively in a public goods interaction a person has to be able to trust that the other person will also contribute. However, this is a mistake. To cooperate requires more than trust in this sense because a person who trusts that another will contribute will still want to contribute zero themselves and free ride on the other person's contribution. Contributing to a public good requires something more than trust in this sense. This is why the experimental evidence is important. Trust is not the same theoretically as cooperation and so we need to be guided by the distinct empirical evidence on effect of group identification on the two types of interaction. There is observational and experimental evidence on group identification weakening trust, but neither on this effect with cooperation. Indeed, the experimental evidence suggests the reverse for cooperation. I have referred to my own experiments that reveal this difference, but it is also found in the meta studies of the experimental evidence in Bailliet/van Lange (2014) and Bailliet/de Dreu/Wu (2014b).

To return to the place of this evidence in Buchanan's argument, group identification is the source of tribalism for Buchanan and he laments the way that contemporary tribalism is adorned by a kind of zero-sum thinking. No one can afford to listen to another group because to do so and so perhaps concede a point

---

3 See, for example, Alesina/Ferrara 2005; Goodhart 2013; Kaufman 2018 and Collier 2018.

is to lose a position in what is a zero-sum world. This is why tribalism is damaging. The evidence of these experiments, however, is that while tribalism in trust games does have this zero-sum character, it does not in public goods type ones. There, the presence of another group does not make life more difficult for the other. To have two groups interacting where there were none, makes one group better-off without injuring the other. This is a positive sum occurrence.

The preliminary lesson, then, or elaboration of Buchanan's argument for 'taking charge of our moral fates', is that we should be encouraging public goods type interactions in societies given to group differentiation and discouraging ones that are akin to trust games. Thus, a society where there are heterogeneous group identifications and in-group biases would do well to attend to shared public goods problems, like those relating to the environment and health, and avoid interactions between members of different groups where a premium is placed on interpersonal trust. Trust, of course, arises in many market exchanges and so this really means such societies should attend to building-up alternatives. Reputations, regulations and contracts are some of the alternatives to trust in many exchanges and so this would mean encouraging vehicles for the public dissemination of information on performance, stronger minimum performance levels being written into regulations and cheaper recourse to formal contract writing and enforcement.

3 Inequality and Cooperation

Granted the conclusion of the previous section on the importance of encouraging public goods type interactions, I focus specifically next on how a different kind of possible tribalism, an economic one between rich and poor, affects behaviour in public goods games. This is because it can nullify the good effects of group differentiation in public goods interactions.

Inequality, typically wealth/endowment inequality, in public goods games has been much studied in the experimental literature and on balance inequality typically, but not always, lowers contributions to the public goods. It is, however, rare in this literature that two separate things associated with introducing inequality are properly distinguished: there are wealth effects as well as positional ones. The point here is that, when a society moves from equality to inequality, holding total endowments constant, the 'poor' under inequality have a smaller endowment and the rich get a bigger endowment than was the case under equality. Thus the behaviour of the 'rich' and the 'poor' might change under inequality for two conceptually distinct reasons. The first is that their wealth/endowment changes and wealth affects behaviour. Second, their relation to others in society changes:

whereas they were all the same in terms of wealth, they are now unequal and this positional change may affect behaviour. This is important not just for the identification of why inequality affects behaviour, it is also important for the identification of a possible rich-poor in-group biases and its origin. I again refer to one of my experiments because it is designed to make this distinction and so allows us to draw insights with respect to this possible economic form of tribalism (i.e. a class based in-group bias).

In Hargreaves Heap/Ramalingam/Stoddard (2016), subjects are randomly sorted into 3 person groups and play a public goods game. In this game the subjects have an endowment and must decide how to allocate it between a private and a public account. The return on any allocation to private account is 1. The contributions to the public account are multiplied by 1.5 and everyone in the group shares equally in this public good (so the marginal personal contribution rate of return is 0.5, providing the incentive to free-ride). This game is repeated 20 times by the same group of 3 subjects and after each round they get information on the total contribution to the public good in the previous round. We have three equality controls: in one everyone has 20, in another 50 and in the final equality control they each have 80 experimental points as an endowment. This is how we will be able to control for the effect of wealth on behaviour in our inequality treatment and so identify the positional effects of inequality (i.e. the fact that with inequality the members of the group have different levels of initial wealth/endowment). In the inequality treatment, each of the 3 members has a different endowment: one has 20, another 50 and the other has 80. The effect of positional inequality, controlling for wealth effects, can now be gauged by comparing how subjects with the same endowment level behave when there is equality and when there is inequality. For example, does the person with a 20 endowment under inequality behave the same as the person with an endowment of 20 when everyone has an endowment of 20? Table 2 makes this comparison for each endowment level.

	20 endowment	50 endowment	80 endowment
Equality	0.44	0.49	0.51
Inequality	0.49	0.50	0.33

Tab. 2: % contribution of endowment to PG

It will be seen that inequality in this sense only has an effect on the behaviour of the 'rich' (those with an 80 endowment level).[4] The 'rich' treat fellow 'rich' people differently to the way they treat others in an unequal group. The 'middle' and 'poor' endowment subjects do not: they treat their fellows in an equal endowment group the same as they do other endowment levels in an unequal group. In this sense, the 'rich' reveal an in-group bias: they are more generous to fellow 'rich' than to people with different and lower endowments to themselves. Further this bias is probably best understood as arising from a form of negative discrimination. The 'rich' under equality are no more generous in % terms than are the 'poor' or 'middle' endowment level equal groups. So they do not respond in this sense any differently to being in their own group than do other endowment levels. But the rich do treat 'others' less generously to create the bias when other endowment levels do not treat their respective 'others' less generously.[5]

There is, thus, a form of economic tribalism in the sense of an in-group bias revealed in this experiment, if only among the 'rich', when there is inequality. Further it seems likely to fuel zero sum thinking. This is because the presence of differentiation in endowment levels lowers the average contribution to the public good. It may give the 'rich' a heightened sense of their own distinction through their in-group bias, but it causes the public good to diminish. The economic tribalism that comes with inequality, therefore, allows a status gain for some (the 'rich') but it comes with a general loss in pay-offs.

To connect this experiment to Buchanan's argument, it supplies an additional and more direct reason for tackling inequality in order to secure his sense of moral progress. In his argument, inequality features as a contributor to tribalism because it provides incentives for the rich to manipulate tribalism in order to secure their positions of power. One thinks, for example, of strategies of 'divide and rule' used by the powerful to weaken the claims of the poor when the poor are heterogeneous in terms of other sources of identity (e.g. race and ethnicity). Indeed, there is experimental evidence that supports the idea that the poor are less likely to vote for redistribution when they have heterogeneous other sources of identification (see Klor/Shayo 2010). The experiment above suggests, however, that quite separately

---

4 In passing, it can also be seen that there is a wealth effect. While the % contribution does not change as endowment levels increase under equality, the absolute contribution level does. This highlights the importance of controlling for wealth when judging the effects of inequality.
5 It is sometimes argued that the withdrawal of the rich arises under inequality not because of any change in public spirit on their part but because these contributions are known to be influenced by reciprocity and the scope for reciprocity is diminished under unequal endowments. There is, however, evidence of a similar effect from inequality in dictator games and this cannot be explained by reciprocity: the rich simply become less pro-social.

from this influence of economic inequality, inequality is itself a source of tribalism with a class based kind of in-group bias that seems to encourage zero-sum thinking. Inequality does not just feed into non-economic sources of tribalism; it creates its own forms of zero-sum thinking tribalism.

As a footnote to this discussion. I have two experiments, Hargreaves Heap/ Ramalingam/ Stoddard (2015) and Hargreaves Heap/ Ramalingam/ Stoddard (forthcoming) where we examine how a competition between teams affects each team's contributions to their own public goods. I mention them now because while there is experimental evidence to support Buchanan's argument that markets as institutions encourage inclusiveness, there is also some more nuanced evidence on the role of competition. This is important because markets can have more or less competition.

It is well known that competition between teams for a prize boosts the teams' contributions to their public goods (e.g. see Erev/Bornstein/Galili1993) and in these two experiments we look at whether this boost is affected by the presence of inter-team and intra-team inequality. In Hargreaves Heap/Ramaligam/Stoddard (2015) we find that when the teams are differently resourced, this competition boost to public goods tends to disappear, particularly among the richer team. This reinforces the insight of the experiment above on the tendency for inequality to impair contributions to public goods and that this occurs because the 'rich' (team in this instance) becomes less public-spirited. The forthcoming experiment looks at teams which are equally resourced but where there is inequality within each team. The inequality within each team matches that in the experiment above: (20, 50, 80) are the endowment allocations for the three members of each team. We put two such unequal teams, like the above experiment, into competition with each other; and we find that the key boost to public good contributions comes from the 'rich' in each team. In other words, the rich, who seem to lose a sense of public spirit when there is inequality within their team, recover most of that public spirit when their team is put in competition with another. It is 'as if' the 'rich' recover their sense of belonging to their team through the competition; whereas in the competition between teams that have different resources, the 'rich' team withdraws from the competition to some degree because the result has become a foregone conclusion.

I draw the conclusion from this footnote, again to supplement Buchanan's prescriptions for moral progress, that the introduction of competition may encourage the positive sum kind of thinking that comes from raised contributions to a public good in an unequal society but only when the inequality within that society is largely within teams and the competition is between teams. If the inequality is more strongly driven by inequality between teams, then competition has the reverse effect.

4 Moral Progress

I turn explicitly in this section to what this evidence on the varieties of tribalism might entail for Buchanan's project of encouraging 'moral progress'.

My first conclusion follows directly from the last section. The case for tackling inequality, acknowledged already by Buchanan, becomes much stronger because it is an independent source of the worrying kind of zero-sum tribalist thinking. This conclusion is reinforced further by the recent arguments of the OECD (2015) and the IMF (2017). They both suggest that at the current time, reducing inequality is likely to promote the growth of output. Or to put this round the other way they find that in many rich countries output growth has been slowing at the same time since the mid 1980s as inequality has been also been growing and the two are related with the result that some poor groups have seen no increase in their living standards and some may have fallen while those at the top have seen theirs power ahead. In these circumstances, there can only be a macro encouragement from the experience of the last 30-40 years to zero-sum thinking because there will be some groups who simply have not experienced living in market society as a project of mutual benefit. A judicious closing of the income gaps directly addresses the experience of zero-sum outcomes of the poor and may succeed in initiating a positive-sum orientation if greater equality also boosts growth, as the OECD and IMF seem to think, and so allows everyone's living standards also to rise.

If tackling inequality becomes central for these reasons, the experimental evidence on the dwindling public-spiritedness of the rich above is not encouraging about the prospects of taxing the rich to secure this reduction in inequality. The point is that a public goods game can be viewed as a voluntary tax game where the tax is productively used and is redistributive because the gains from this social production are equally shared. From this perspective, what the public goods game experiment under equality and inequality tell us is that the rich become less willing to tax themselves for redistributive purposes when there is inequality. In short, the rich become less likely to volunteer to pay higher redistributive taxes exactly when this is what is needed to combat Buchanan's worrying zero-sum forms of tribalism. This puts at the top of Buchanan's agenda for moral progress, for me at least, the following question: how to encourage, as it were, a form of enlightened self interest in redistribution among today's rich and powerful?

I have no proper answer to this question, but it becomes doubly important when connected with the first experimental insight about varieties of tribalism: i.e. public goods game-like interactions, and unlike trust ones, can encourage positive sum thinking. Inequality in public goods games, in effect, forecloses on this possibility—or so the experimental evidence suggests.

There is, though, one thought regarding what might be an element of an answer to this question that comes from the experimental literature. That is, our perception of the role of luck in outcomes affects our apparent willingness to redistribute. There are several experiments where the same level of inequality is generated either by luck or by individual decision making or effort and they find that subjects engage in more redistribution in the former than the latter (e.g. see Cappelen et al. 2013). In other words, people seem to be willing to redistribute more when the outcomes are thought to arise from luck than individual effort/decision (i.e. 'merit'). Thus part of an answer might be to change beliefs particularly among the rich over how much luck is responsible for their wealth.

Of course, this will be difficult if current beliefs about the respective roles of luck and merit are right. Beliefs are well-known generally to have a self-serving character and so the rich seem bound to incline to the merit side of the explanation, while the poor will find the luck side more cognitively satisfying. Both cannot be right, but what matters for moral progress, on the basis of the argument here, is getting the rich to believe that luck plays a bigger role than they are naturally inclined to think and what hope of doing this unless this is actually the case?

The strongest argument, I suggest, for thinking that luck plays the upper hand, and increasingly so, comes from Hayek. He may have thought that redistribution in the name of the 'mirage of social justice' was an obvious mistake and so he is not the most obvious ally in a project for redistribution, but he provides one of the most powerful arguments for why economic outcomes are increasingly driven by luck. This is his fundamental message or insight when arguing that use of knowledge is the economic problem (e.g. see Hayek 1945). The counterpart to his problem of knowledge is that our beliefs about the world and how action connects with outcomes are necessarily imperfect. Outcomes are often not foreseeable in even a probabilistic sense and so market outcomes do not arise simply from individual *rational* calculations. They arise from individual decision but not from individual design. In short, from the perspective of the necessarily imperfectly informed individual, outcomes must appear as driven in some degree by luck, that unforeseeable component.

> The market leaves the particular combination of goods and its distribution among individuals, largely to unforeseeable circumstances –and in this sense to accident. It is, as Adam Smith already understood, as if we agreed to play a game, partly of skill and partly of chance. (Hayek 2014, 310)

What gives luck the increasing upper hand is that Hayek argues, plausibly, that as an economy grows, so does its complexity, and with this the knowledge problem.

> The more men know, the smaller the share of all that knowledge becomes that any one mind
> can absorb. The more civilized we become, the more relatively ignorant must each individual
> be of the facts on which the working of civilization depends. The very division of knowledge
> increases the necessary ignorance of the individual of most of this knowledge. (Hayek 1960,
> 78)

Thus, my second conclusion is, baldly, that the 'rich' need to read Hayek. The
thought is that if more people perceive the world as uncertain and risky, they are
more likely to embrace forms of insurance. But in the absence of easy identification
of the roles of luck and merit in any outcome, social insurance will have advantages
over private insurance, particularly when there is asymmetric information. Thus an
appreciation of Hayek's knowledge problem will incline people to embrace further
forms of social insurance. It also means there will likely be redistribution as this is
what typically happens with social insurance because everyone pays a premium
and only those who suffer adversity receive a benefit. Crucially, though, this is not
why one might expect the rich to sign-up. They will vote for insurance when they
perceive that it is an arrangement from which they (and others) will benefit. This
is always possible because this is the magic of pooling risk when individuals are
risk averse: it is a positive-sum social innovation.

Of course, this strategy turns redistribution into a by-product of a desire for
insurance. Prime the desire for insurance because the world is increasingly subject
to uncertainty and risk and you get redistribution via the backdoor. It is not a
proposal that directly addresses how to activate a sense of fairness in support of
redistribution. This may be a weakness because there is no direct motivational
reason for redistribution. But its strength is that it plays into what is known from
laboratory experiments about people's willingness to redistribute. The point is
that we also know from these experiments that people are more or less likely to
support redistribution depending on their sense of what is fair. It is not just a
matter of how important luck in determining outcomes is perceived to be. The
difficulty however, with pushing directly the fairness route to redistribution is that
significant individual differences are also revealed in these experiments about what
is considered fair. Fairness is naturally contested, in other words; whereas who
can doubt, post COVID-19 and the climate emergency, that the world is becoming a
riskier place?

5 Conclusion

In support of Buchanan's agenda of moral progress (i.e. tilting our moral compasses
towards inclusivity and away from 'them' and 'us' zero-sum tribalism), I have

argued that we need a more complicated understanding of tribalism than Buchanan gives us. First, not all kinds of group identity differentiated interactions breed 'them' and 'us' zero-sum dynamics. For instance, group affiliations in public goods interactions are positive-sum in their effects (absent inequality, of course). Second, economic inequality does not just exacerbate identity based zero-sum tribalisms, it is a form of economic tribalism with its own potential for distinct zero-sum dynamics (notably in the otherwise helpful public goods interactions).

These arguments are based on laboratory experiments where subjects are either artificially affiliated with different groups or given equal/unequal endowments. They depend for their practical value, therefore, on such laboratory experiments having a measure of external validity. I have given some reasons for thinking this might be the case. Granted that these reasons are sufficiently persuasive, the evidence makes inequality a key ingredient in the generation of zero-sum tribal dynamics. Tackling inequality is, therefore, I suggest the top item on the agenda for moral progress. The fact, however, that the tribalism of inequality is most noticeable in these experiments among the rich and it is the rich who, of course, would lose out when tackling inequality makes this item on the agenda of moral progress a tough political one. I have speculatively suggested, nevertheless, that Hayek might help here. If we appreciate along with Hayek that the world is increasingly uncertain, we may be more inclined to sign up to new forms of social insurance to cover such uncertainties; and social insurance is a wonderful positive-sum social innovation that just happens to be redistributive.

References

Alesina, A./E. La Ferrara (2005), Ethnic Diversity and Economic Performance, in: *Journal of Economic Literature* 43(3), 762–800

Bailliet, D./P. Van Lange (2014), Trust, Conflict and Cooperation: A Meta-Analysis, in: *Psychologic Bulletin* 139(50), 1090-1112

— /C. De Dreu/J. Wu (2014), Ingroup Favoritism in Cooperation: A Meta-Analysis, in: *Psychology Bulletin* 140(6), 1556-1581

Buchanan, A. (2020), *Our Moral Fate. Evolution and the Escape from Moral Tribalism*, Cambridge/MA

Cappelen, A./J.Konow/E.Sorensen/B.Tungodden (2013), Just or Luck: An Experimental Study of Risk Taking and Fairness, in: *American Economic Review* 103(4), 1398–1413

Collier, P. (2018), *The Future of Capitalism*, London

Corr, P. J/S.Hargreaves Heap/C. Seger/K. Tsutsui (2015), An Experiment on Individual Parochial Altruism Revealing no Connection Between Individual Altruism and Individual Parochialism, in: *Frontiers in Psychology* 6, 1261

Erev, I./G. Bornstein/R. Galili (1993), Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment, in: *Journal of Experimental Social Psychology* 29(6), 463–478

Goodhart, D. (2013), *The British Dream: Successes and Failures of Post-War Immigration*, London

Hargreaves Heap. S./D. Zizzo (2009), The Value of Groups, in: *American Economic Review 99*, 295-323

— /A. Ramalingam/B. Stoddard (2015), Doggedness or Disengagement? An Experiment on the Effect of Inequality in Endowment on Behaviour in Team Competitions', in: *Journal of Economic Behavior and Organisation* 120, 80–93

— /B. Stoddard/A Ramalingam (2016), Endowment Inequality in Public Goods Games: A Reexamination, in: *Economic Letters* 2016, 146, 4-7

— /A. Ramalingam/B. Stoddard (forthcoming), Team Competition When there is Within Team Inequality, in: *Journal of Behavioral and Experimental Economics*

Hayek F. A. (1945), The Use of Knowledge in Society, in: *American Economic Review* 35(4), 519-530

— (2014), Competition as a Discovery Procedure, in: *Collected Works of F.A. Hayek*, volume 15, *The Market and Other Orders*, edited B. Caldwell, Chicago

— (1060), *The Constitution of Liberty*, Chicago

IMF (2017), Tackling Inequality, in: *IMF Fiscal Monitor*, October

Kaufman, E. (2018), *Whiteshift: Populism, Immigration and the Future of White Majorities*, London

Klor, E./M. Shayo (2010), Social Identity and Preferences over Redistribution, in: *Journal of Public Economics* 94: 269-278

OECD (2015), *In It Together: Why Less Inequality Benefits All*, OECD Publishing

Putnam, R. (2007), E pluribus unum: Diversity and Community in the Twenty First Century, in: *Scandinavian Journal of Political Science* 30(2), 137-174