Social Macroeconomics
Working Paper Series

Diverging identities: a model of class formation

SM-WP-2019
January 2019

Paul Collier, Blavatnik School of Government,
University of Oxford

Diverging identities: a model of class formation

Paul Collier

Second Revision, January 2019

Blavatnik School of Government,
Oxford University

ABSTRACT

This paper is an application of Identity Economics. Since the literature in this field is recent, the paper begins with an extensive review of the key contributions. The current paper analyses the process and psychological costs of social polarization arising from economic inequalities. It may have some application to the current social divisions evident in to the votes for Brexit and Donald Trump and protest movements such as the *gilets jaunes*. In a simple model, people rationally maximize their utility from esteem, by selecting a subjective *salient* identity from two objective identities: nationality and job. The model shows how an increase in wages for the upper half of the population can lead those with high incomes to drop nationality as their salient identity, forming a new 'elite' class. This rational switch in the identity of high-income workers has both efficiency and redistributive effects, reducing aggregate utility and generating regressive transfers.

Diverging Identities: a Model of Class Formation[1]

> 'Actually there were only two forms of existence, I reflected: one that was tied to place and one that wasn't.'
>
> Karl Ove Knausgaard, *Some Rain Must Fall* (*My Struggle*, Vol 5)

1. Introduction

It is now a commonplace that people are becoming polarised into rival group identities to which they attach subjective importance, and which are believed to generate distinct behaviours (Chua, 2018, Goodhart, 2017, Williams, 2017). This is usually explained by invoking some dichotomous objective characteristic such as the level of education, or location, which objectively determines these rival subjective identities, and which directly explains why they are deemed important by those who hold them. However, such objective characteristics are not new, and in a previous era were not assigned much subjective significance. In this paper I apply and extend recent economic research on group identity to better understand what might be happening.

Economic research to incorporate group identity into behaviour, a line of work commonly termed 'Identity Economics', was pioneered by Akerlof and Kranton (2000). In their paper, subjective identification with a group directly entered the utility function and affected behaviour through the influence of group norms. Since 2000, the research has developed into two distinct branches: the processes by which subjective group identity is acquired; and the various channels by which, once acquired, it can influence the behaviour and utility of group members. While the agenda falls within Behavioural Economics, it is a distinct departure: its focus is on groups rather than individuals and, by definition, the behavioural effects it investigates vary between groups and so cannot be universal traits explicable by a socio-biological evolutionary process. They are in some sense 'cultural', albeit that in ordinary usage the term has wider connotations.[2]

The present paper shows how a society can become bifurcated into two subjectively significant 'classes' by a small change in the range of one continuous objective variable, (which can be thought of as income). I model a modest increase in income inequality that results in a rational subjective process of bifurcation into class identities. In turn, because these new identities affect utility, the change redistributes utility and can reduce it in aggregate. By demonstrating these consequences for utility, the model enhances our understanding of why longstanding objective characteristics have acquired new popular significance.

[1] I would like to thank Tim Besley and three referees for comments on a previous draft, and Michael Blake for research assistance.

[2] An association of economists, ERINN (Economic Research on Identity, Narratives and Norms), reflects this recent body of research.

The paper also extends the research on Identity Economics by clarifying and incorporating two distinct types of subjectively important identity, both of which have featured in the literature, and both of which can confer utility. Some subjective identities are directly based on objective characteristics, such as 'wage-earner'. Other subjective identities are defined not by the objective characteristics themselves, but by the observed difference in choices of salience that people make as between them. Thus, everyone has a job, and everyone lives in the same country, but if some choose to make their job salient and others choose to make their country salient, the society has divided into two new identity groups defined by these choices. As I will show, this distinction matters.

The paper proceeds as follows. In Section 2 I review the pertinent literature and relate it to the present paper. In the following three sections I set out the model and derive the results of an exogenous change in the level and distribution of wages. In Section 6 I discuss possible extensions, including those suggested by features of some existing models that have been flagged in Section 2. Section 7 concludes.

2. The recent literature on group identity

A core idea of Identity Economics is that some objectively given identities of economic actors may generate utility for them. The mechanism can be direct, from a sense of belonging to a group. It may also be indirect, through two distinct channels. One is the esteem that may be conferred on the individual by others in the group. Typically, groups develop their own norms of behaviour, and by adapting behaviour so as to conform to these norms, individual members can generate esteem from the other members of the group. The other channel is that membership of the group may confer esteem on all its members, bestowed by non-members. For example, a group may be regarded as prestigious by the entire society.

This expansion of the utility function to include belonging and esteem is as securely grounded in socio-biology as is the desire for consumption. The neurological instinct for the urge to belong is generated by the release of oxytocin. The original evolutionary advantage conferred by the release of oxytocin was to bond parents to their children, but this gradually became co-opted for the larger purpose of cooperation within a group (MacDonald and MacDonald, 2010). People naturally tend to identify with those other people with whom they share some similarities. The neurological instinct to seek the esteem of others is generated by the release of testosterone. Our sensitivity to esteem is acute: when humans meet, we detect differences in social rank within $1/25^{th}$ of a second (Sapolsky, 2017).

In the pioneering model of Akerlof and Kranton (2000), an identity was an attachment to a group that directly conferred utility. The group had norms, so that in addition to valuing belonging, behaviour that conformed to these norms generated utility through esteem. Since different groups may be awarded different levels of esteem by non-members, this introduces a potential tension between the urge to belong, and the urge for esteem. In order to get esteem, people may be willing to attach themselves to prestigious groups with which they have less in common that less prestigious groups.

An important precursor to Akerlof and Kranton was Rotemberg (1994), who demonstrated the potential for the endogenous emergence of reciprocal altruism among rational, initially self-interested individuals. He pointed to a key problem in the emergence of such mutually beneficial reciprocal altruism; namely how a self-interested actor could credibly establish his altruism towards another actor. Rotemberg considered signalling actions such as observable and hard-to-fake body-language, and costly gifts. At the end of his article he even pointed to social networks as a neglected area for economic research on reciprocal altruism. In retrospect, he was reaching towards the missing concept of subjectively chosen identity. While such an identity can be abandoned, it is analogous to an investment and so abandonment is costly (Benabou and Tirole, 2011). In effect, that act of choosing an identity is, among other things, the commit technology that Rotemberg was seeking. Consequently, shared subjective identity tends to predispose members towards each other.

In one of the first subsequent models, Bisin and Verdier (2001) focused on the process by which a group identity was acquired. Their model introduced two cultural identities, coexisting in the same society. Abstracting from esteem, they focused exclusively upon the direct contribution of identity to utility, studying its transmission between generations. Parents have one of two cultures. They get utility not only from their own identity but from the identity that their children adopt; crucially getting more utility if their children identify with the culture that they themselves have adopted. Children acquire their culture neither genetically, nor through a conscious rational choice, but through social interaction. This assumption is consistent with the social psychology literature which suggests that the capacity for rational thought does not develop until around the age of 14, whereas group identity is established earlier (Hood, 2014). The social interactions that Bisin and Verdier assume set identity are partly with their parents ('direct' transmission), and partly with other members of society ('oblique' transmission). Parents can spend resources on direct transmission, for example by occupying their children's time by playing with them, rather than letting them watch television. Parents can also spend resources on oblique transmission, for example by buying a house in a catchment area that gives them access to a school that has children from their own culture. If parents do not do this, children still acquire a culture, simply through costless social interaction with other people. The culture they adopt as a result of social interactions simply reflects the cultures of those with whom they interact, this being determined by a random draw. This form of acculturation is costless and so, in effect, a public good.

Their key result is to show that if the two cultures both persist in the social equilibrium, (which they show is a feasible outcome), parents maximize their own utility in a way that reduces wellbeing in the society. Specifically, parents from each culture are driven to devote resources to direct and oblique cultural transmission, rather than leaving acculturation to the costless process provided by the public good. Some of this is a zero-sum game, since these parental efforts reduce the efficacy of the costless process for parents belonging to the other culture.

The Bisin-Verdier model is directly pertinent for the present paper. As in that model, I posit two co-existent cultural identities. Although in its basic form the new model does not include behaviour directed towards the inter-generational transmission of identity, it readily lends itself to this extension. Once inter-generational transmission is incorporated into it, it

generates a highly specific and testable prediction about parental behaviour. In Section 6 I present clear quantitative evidence from Britain and America that is consistent with this prediction. Specifically, I show that in response to the wage shock that the present model analyses, within the newly emerging minority identity group there has indeed been a substantial increase in family resources devoted to both direct and oblique transmission. I outline the implications of this new behaviour for an additional layer of adverse effects on aggregate utility and its distribution. A further implication of their model, though one that they do not discuss, is that this social waste will be at its peak, *ceteris paribus*, if the society is equally split between the two cultures. This will also be considered in Section 6.

Chandra (2012) provides a considerably richer conceptual analysis of the evolution in group identities. Although her focus is on ethnic identity, which is not the subject of the present paper, the conceptual apparatus is more general. She and her co-authors make a fundamental distinction between 'attributes' and 'categories'. The former are the objective characteristics which are the raw materials from which identity groups may be formed; the latter are subsets of these characteristics which are given psychological significance in a particular social context: they are the identity groups constructed by social entrepreneurs from these raw materials. Such a change in group identity for a given set of attributes is the subject of this paper. I model the choice of which of two attributes to make salient: in the process, those who make these choices divide into two new, socially created categories.

Chandra considers five mechanisms by which a given set of attributes can be rearranged into changes in categories. One of the five mechanisms, 'passing', is particularly pertinent. By 'passing', Chandra refers to emphasis upon an attribute which, since it is also held by the members of some other group, permits someone who is initially part of a different group to be accepted as a member by that group: an example of its use familiar in America would be 'passing for white'. Chandra's concept of 'passing' is pertinent, because it constrains the scope for rational actors to deceive themselves. They cannot internalize the idea that they are members of a group that rejects them as members. In the model presented below, the most highly-skilled wage-earners create a prestigious new category, and deemphasize their previous inclusive identity. Their rules of acceptance into the group (which may be implicit, but are nevertheless well understood) are that the choice of salience is switched (from place to job), and that any member of the new group must be earning more than any non-member. Those who continue to make place-based identity salient cannot fool themselves that they share salient identity with this group: they can no longer 'pass' for members of the same salience category. I impose this rationality constraint on the adoption of an identity in the model presented below.

Whereas Bisin and Verdier abstracted from esteem, two more recent models of identity make it central. Each sets the context as the formation of identity in school, which it models not as a random process of social interaction, (as in the Bisin-Verdier model), but as a rational choice of esteem-seeking behaviour. This is not a challenge to the assumptions of Bisin and Verdier. People hold multiple identities, each subjectively significant in a social context. Bisin and Verdier model the primary acquisition of cultural identity, which is formed prior to rationality, while the two models considered below focus on identities which carry differential prestige and so are reasonably regarded as the result of rational choice under constraints. For purposes of analytic tractability, each model again considers

only two identities. Eguia (2017) starts from the assumption that one identity is more prestigious than the other. This is a feature that was not pertinent for the Bisin and Verdier model, in which each cultural group values its own identity symmetrically more than that of the other group. In the Eguia model, some children come to the school with an elite identity acquired from their parents, while others come to it with the inferior identity, similarly acquired from their parents. Non-elite children wish to switch to the identity of elite children, and can do so, but only if elite children accept them as having elite identity. In the language of Chandra, they have to be able to 'pass'. I have emphasized this condition for membership of the elite group, for reasons that will become apparent when I discuss the model presented in this paper. Children in the elite group admit those from the non-elite group selectively, screening them according to academic attainment. In turn, the academic performance of non-elite students is affected by the effort that they put into studying, this being an observable behaviour. Defection by high-effort members of the non-elite group further reduces the esteem generated by membership of that group. As a defensive reaction, those members of the non-elite group who are unwilling or unable to signal the required performance, punish would-be defectors: the behaviour sometimes stigmatised by non-elite African American children as 'acting white'. As in the Bisin-Verdier model, these privately rational behaviours – the screening adopted by the elite group, the signalling of would-be entrants to the group, and their punishment by remaining members of the non-elite group – can in aggregate be socially costly.

In the present model, I adopt key features of the Eguia model. Members of the non-elite group wish to join the elite group, but to be accepted must make an observable sacrifice which only those best-placed to be members of the elite are rationally prepared to do. The choice is rational in two distinct senses: it is determined by utility maximization, and it excludes the adoption of the identity by those who do not 'pass'.

In a further model of the classroom, Robert Akerlof (2017) investigates a different response to exclusion from an elite group: the creation of a rival prestigious identity. Initially, all the children are in the same group, 'Nerds', whose norm is academic success. Reflecting the norm, they reward each other with esteem in proportion to academic success. As in the Eguia model, this is the result of effort and innate ability. But the children around the bottom of this hierarchy can choose to reject this identity and adopt another one in which they are able to be more successful: 'Rockstar'. A sub-group of the class rationally defects from the 'Nerd' identity, and judges its members by a different set of criteria, thereby generating more esteem, and hence more utility. As in all models of identity, this privately rational esteem-seeking behaviour has consequences for both aggregate social welfare and its distribution that need not be benign.

In the model presented in this paper, I adopt from the Akerlof model this option of exiting an initially common identity to create a new identity that is superior for the group. Whereas in the above model it is those with the lowest esteem who exit, in the one presented below it is those who initially already enjoy the highest esteem. However, they exit for the same reason: by doing so they generate yet higher esteem.

The model that is closest to the model presented here, in that it has a similar context – a choice of identity between job and nation - is Shayo (2009). However, the assumptions,

focus and behaviour analysed in the two models are interestingly different. Indeed, the two models lead to very different results. As in the present model, actors make a choice as to which of two objective identities they will make subjectively salient. In the Shayo model this is nation or 'class'. Objectively, everyone is a member of the same nation, and one of two 'classes' – a minority elite class, or a majority non-elite class. In contrast to the present model in which the critical choice of identity is made by high-wage earners, in the Shayo model, it is made by low-wage earners. This is because the core of that model is the effect of public policy as determined by the preferences of the median voter. In the Shayo model there is a structural cliff in the wage distribution: all members of the elite class receive the same high wage, and all members of the non-elite class receive the same low wage. Hence, the two class identities are objective facts defined by the height of the wage cliff. Within each class, all members are identical so that the key decision variable, which is the choice of salient identity of the non-elite class, yields all-or-nothing outcomes: at some critical threshold of the wage cliff all members of the non-elite class switch their salient identity together. This exogenously imposed two-class wage distribution also contrasts with the present model in which there is a continuum of wages across the society, and the size of each class is determined endogenously.

If members of the non-elite opt to make their nationality salient, then they receive the esteem associated with the nation. If they opt to make their class salient they receive the esteem associated with their wage relative to that of the elite class. Finally, and critically, there is a feedback from these choices of identity to political outcomes. The setting is a democracy in which public policy of redistribution is set by the median voter, which given the assumptions is the non-elite class. This creates a feedback from the choice of identity of the median voter, who by assumption is a member of the low-wage class, onto the wage distribution. It is also why the behaviour of the elite class is of no consequence, although consistent with the present model, those earning high wages are less likely to make their nationality salient than those earning low wages. By assumption, if the non-elite chooses to identify with the nation, the associated norms are less oppositional towards the elite than if it chooses to identify with its class, where the norms are more aggressively redistributive. The exogenous variation in the Shayo model is the initial size of the wage cliff, and its key idea is that there can be multiple equilibria. With a wide wage cliff, the esteem from making class salient rather than nation is lower than if the wage cliff is narrow. But if the non-elite class makes nation salient then the wage cliff will indeed be wider because public policy will be less redistributive. Hence, it is possible, depending upon how rapidly class esteem declines as a function of a wider wage cliff, for there to be two locally stable equilibria: a narrow wage differential, class identity, and strong redistribution, and a wide wage differential, national identity, and weak redistribution. By construction, the utility that the non-elite gets from income is lower in the latter equilibrium, but this is offset by the esteem it gets from believing that it shares identify with the high-wage group. Shayo then presents some data suggesting that this outcome of multiple equilibria is common.

As I will show, with superficially similar assumptions, the present model generates very different results. A key change of assumption is to impose a constraint upon rational beliefs about identity that has become common to the literature since Chandra (2012), but which was not adopted in the earlier Shayo model. This, and other differences, are best discussed as the building blocks of the new model are presented.

Finally, the model of Besley (2016) fuses features of both the Bisin-Verdier and Shayo models, combining inter-generational transmission of culture with a feedback onto political outcomes that generates multiple equilibria. As with the other models it presents a dichotomous choice: people hold one of two political values, X or Y. Initially, both beliefs co-exist in the society in arbitrary proportions. People mate according to affinity of political belief and their children acquire their own beliefs from those of their parents. Occasionally, however, akin to Bisin-Verdier, a belief mismatch occurs in a marriage. The child of X-Y parents then adopts the beliefs of whichever parent is the happier. The happiness of each parent depends upon whether their own beliefs coincide with those of the majority, since these will be the ones implemented in public policy. If the X value is initially in a small majority, the polity will adopt X-friendly policies, and so in X-Y marriages the X parent will be the happier. In consequence, the children of such marriages will themselves adopt X values, and so over time the X majority will grow larger. By entirely symmetrical reasoning. If the Y value is initially in a small majority, it too will grow larger over time: two such societies will diverge over time, amplifying the differences in their values.

In the present paper, the setting is a citizen in a society such as Britain or the USA. As with the other models, I consider a binary choice as to which of two objective identities – nationality and job – the citizen should choose to elevate by making it subjectively salient. I keep the set-up skeletal in order to bring out the implications as straightforwardly as possible.

3. The Set-Up

All actors each have two objective identities, their nationality and their job. The all have the same nationality, N. They all have a job, but the income generated by the job differs. Actors get esteem from each identity, and this esteem generates utility: the model abstracts from all other sources of utility. Thus far, the model is structurally similar to that of Shayo (2009). However, unlike the Shayo model, Each actor gets four distinct contributions to self-esteem from these two objective identities.

The first source of esteem is from national identity. National identity confers the same amount of esteem on each actor, denoted by the amount N.

The second source of esteem is from the job. The job confers a different amount of esteem on each actor, the variation depending upon their position in the distribution of wages. In contrast to the Shayo model with its assumption of only two wage rates, I consider a continuous distribution of wages. This enables the model to have a marginal actor who is indifferent between the two possible choices of identity, so that the size of each identity group can be determined endogenously, rather than being exogenously imposed by the wage structure as in the Shayo model. For tractability, I specify this distribution as uniform. Without loss of generality, I specify the esteem-utility generated by the wage, U_{Wi}, as being linear in this wage ranking, minus a constant. The constant is set, for convenience, such that the lowest-ranked wage earner gets zero esteem and the highest ranked gets W, with the median earner getting 0.5W:

$$U_{Wi}, = Wr_i \quad (1)$$

Where W denotes the utility generated by the highest wage, and r_i is unity for the highest-ranked wage, zero for the lowest-ranked, and linearly interpolated between them.

An attractive feature of this specification is that esteem is not assumed to be a zero-sum game, but rather comes from the absolute level of achievement. In this case, the higher is productivity, and hence the wage, the higher is esteem. The expectation is therefore that an increase in productivity will increase aggregate wellbeing.

In addition to these objectively given sources of esteem, the actor has the scope for generating further utility by choosing to bestow *salience* upon one or other of the objective identities. That is, the actor can regard herself as first-and-foremost defined by her job, or by her nationality. As is apparent from the previous section, this move from objective characteristics to a choice of subjective identity is standard in models of Identity Economics. For tractability, I specify the effect of bestowing salience on a characteristic: whichever identity that the actor chooses to make salient doubles the potency of that identity, and so doubles the amount of utility generated by it. Consistent with individual rationality, in making this choice, the actor is assumed to maximise utility.

It might seem that this choice is a simple matter, with nationality being chose if and only if $N \geq Wr_i$. However, in making this choice the actor generates a third identity, namely membership of the group of people who have made the same choice of identity. This is, in effect, a rudimentary form of subjective class formation. If all actors make the same choice of identity no such class formation occurs, but if some choose nationality and others choose job, then by joining one or the other of these two classes the actor opens up a fourth source of utility, namely that conferred by membership of the group. For simplicity, the model assumes that the esteem generated by this group identity reflects the average within the group of the sum of the other three sources of esteem: nationality, job, and the boost to nationality or job bestowed by salience. This feature of the model is a significant innovation that contrasts with the Shayo model. There, if the structurally determined low-wage class chooses to make nationality salient, then, as in the present model, it receives utility N, but even if the high-wage class differentiates its identity from the low-wage class, (which it has a strong incentive to do), the low-wage class does not notice. Since the work of Chandra (2012) on identity and 'passing', subsequent scholars have imposed a rationality restriction on the adoption of an identity: essentially, people cannot choose to identity with a group that requires characteristics that they lack, and which consequently rejects them as members. The Shayo model rests on an assumption that does not satisfy this restriction. The justification for the restriction is evident from societies such as present-day Britain and the USA, where the mutual polarisation of subjective identities is unmissable, and as with other group identities, these groups potentially bestow different amounts of esteem that reflect their different compositions. This feature, that different choices of salience between objective characteristics create a new objective characteristic, and thereby new groups, to my knowledge is new to the literature. Combined with the endogeneity of group size, replacing that exogenously imposed by the assumption of a structural wage cliff, the model

generates very different results from those of the Shayo model, despite the superficially similar characterisation.

4. Class Formation

As set up, everything is determined by the relative values of W and N. Whether the society is homogenous or divides into two classes depends upon the existence of a critical actor, c. This *c*-th actor is defined as being indifferent between making her nationality and her job salient. If there is a critical actor then all actors in more prestigious jobs, for whom $r_i > r_c$, will make their job salient, and all actors in less prestigious jobs, for whom $r_i < r_c$, will make their nationality salient. Recall from (1) that W denotes the productivity and utility generated by the highest-earning worker, and that r_i is the relative productivity applicable to the *i*-th worker, so that Wr_c denotes the productivity of this critical actor.

The c-th actor faces the following choice.

If nationality is made salient then utility will be generated from the following four sources:

The objective component of national identity, N
The objective component of her job, Wr_c
The boost conferred directly by salience is N
The esteem generated by membership of the class of those who make nationality salient is the average of its three components:

$$\{2N + (Wr_c/2)\}/3 \qquad (2)$$

The term in (.) denotes the contribution of average job esteem generated in the group, which is uniformly distributed on the range from $W.r_c$ to zero.

So that total esteem is:

$$2N + Wr_c + \{2N + (Wr_c/2)\}/3 \qquad (3)$$

If, instead, the c-th actor makes her job salient, then utility from the four sources will be generated as follows:

The objective component of national identity, N
The objective component of her job, Wr_c
The boost conferred directly by salience is Wr_c
The esteem generated by membership of the class of those who make their job salient is:

$$[N + W + Wr_c]/3 \qquad (4)$$

So that total esteem is:

$$N + 2Wr_c + [N + W + Wr_c]/3 \qquad (5)$$

Since the critical actor is indifferent, these four components must sum to the same amount for each choice. Hence:

$2N + Wr_c + [2N + (W.r_c/2)]/3 = N + 2Wr_c + [N + W + Wr_c)]/3$ (6)

Rearranging:

$r_c = [(8N/W) – 2]/7$ (7)

Consider the situation in which the esteem from national identity is so high that even that from the highest remunerated job only just equals it, so that W = N. Even in this case the society divides into two classes. On the specific numbers, $r_c = 6/7$, so that the top-earning seventh of the society chooses to make their job their salient identity. In making this choice all but the top earning worker actually get less esteem from their job than from their national identity and so their choice directly generates an avoidable average loss. For the average worker making this choice, the loss is the simple average of the N/7 loss of the critical actor, and the breakeven of the most highly paid worker: hence, it is N/14. Yet the choice is rational because, by identifying with the elite class, they get a larger compensating gain. But both the offset loss and the net gain are entirely at the expense of those who do not change their salient identity.

For the critical actor, since by definition she makes no compensating gain in esteem from switching class, this loss is N/7 as before. Since the loss of esteem is the same for all actors in the class, this is the loss for each of them. Summing the consequences, for 6/7ths of the population there is a per capita loss of N/7, whereas for one seventh of the population there is a gain of N/14. Hence, there is a per capita average net loss of 11N/98, or approximately N/9. Were we to switch from Utilitarian to Rawlsian ethics in which the society is judged by the circumstances of the least advantaged group, the welfare loss would be judged far more serious because the losses are being borne exclusively by this group. This is an inefficient transfer from the disadvantaged to the advantaged. In contrast to the Shayo model, it is driven not by the ‘false consciousness’ of the low-wage majority in choosing to make nationality salient, but by the entirely rational, self-serving decision of the highest wage earners in abandoning their national identity in favour of making their job salient. A corollary is that *from the perspective of the elite class*, the low-wage class is now indeed distinctively ‘nationalistic’.

Note that the potential tension between the psychology of belonging – the desire to identify with those similar to oneself; and the psychology of esteem – the desire to associate with those better than oneself, is finessed. As in the Eguia model, those who join the job-salient group are as similar to each other as possible: they are defined by their high rank. The critical (indifferent) actor is equally similar to her neighbours in the ranking, each of whom rationally opt for different groups.

5. Comparative Statics

Having seen the simple mechanics of the model, I now apply it to two types of social change, using comparative statics. The first is the consequences of a decline in the objective

esteem generated by identifying with the nation, the second is a rise in productivity and wages for the upper half of the workforce.

A decline in national prestige

The prestige of a nation can change: it might win or lose a war; or gain or lose an empire. For example, in the USA, the post-1945 generation could take pride in a massive military victory, whereas the post-1968 generation was embarrassed by mounting military defeat in Vietnam. This change can be represented by a decline in the value of N. The previous analysis readily adapts to portray the comparative statics of such a situation. If the value of N is initially $\geq 9/8$ then there is no class formation: everyone chooses to make national identity salient. We have already seen that if it drops from this value to unity, class formation occurs. The only addition introduced by the comparative statics is that there is a loss of esteem for everyone of $2\Delta N$, which is then reduced for those who switch salience, and compounded for those who do not, each by the redistributions already discussed.

An increase in the productivity of high wage earners

Reverting to national prestige as a constant, I now consider the consequences of an objective increase in wage inequality such as has occurred in most OECD societies during the past 40 years. I begin from a situation in which wage inequality is sufficiently modest that the society is cohesive: everyone chooses to make their nationality their salient identity. Given the parameters of the model this occurs as long as $N \geq 9/8$, and for specificity I assume that this condition holds as an equality.

Now suppose that wage inequality, and the dispersion of esteem associated with the job, increases. To mimic the increased wage inequality that has been common, while retaining the simplicity of the model, I assume that below median income, wages remain unaltered. Above the median, wages increase in proportion to the excess of income over median income: specifically, I will assume that this premium over the median doubles. This is a crude characterisation of the stylized facts: median income has stagnated, while wages above the mean have increased substantially. The specificity of the example enables us to generate precise consequences for each of four different groups in the society, showing both the overall change in efficiency, (the absolute amount of wellbeing in the society), and its distribution. The price that is paid is merely some tedious arithmetic.

Recalling that in this social change, wages only increase for those above the median, for all those whose incomes increase, the esteem generated directly from the new higher wage, expressed relative to the former highest wage, W, is now given by:

$$W(2r_i - ½) \qquad (8)$$

Now that the wage premium for those above-median income has doubled, for 18.25 percent of the workforce $W(2r_i - ½) \geq N$ and so they have a direct incentive to switch their salient identity. Beyond this point, switchers take a direct hit, which for the critical actor will be $N - W(2r_{c.} - ½)$. For the switch to be rational, this must be compensated by an offsetting gain from the difference in esteem between the two classes.

If the critical actor chooses to make nationality salient, (and $1 > r_c > ½$), total esteem is:

$2N + (2Wr_c – ½) + [2N + 9r_c/8 – ½W]3$ (9)

which simplifies to:

$9N/3 + 19Wr_c/8 – 2W/3$ (10)

If instead, the critical actor chooses to make the job salient, total esteem is (after simplification):

$4N/3 + 14Wr_c/3 - 2W/3$ (11)

Setting (10) = (11), (the equivalent of (6) above), and solving:

$r_c = (32/55).(N/W).$ (12)

Normalising on W and recalling that N = 9W/8, this yields $r_c = 0.6545$. Hence, overall, slightly over a third of the population, 34.55 per cent, now makes their job their salient identity. Of these, 16.3 per cent of the population are switching despite directly gaining less esteem from their job, even with its higher productivity, than they get from their national identity. They switch because of the greater esteem from being associated with the group that chooses to make their job salient.

The increase in productivity produces a direct gain in esteem, and indirect effects from the changes in the choice of salience. One attractive feature of the model is that esteem is not assumed to be a zero-sum game: if a worker becomes more productive, her esteem goes up correspondingly, and there is no counterpart direct loss of esteem inflicted on workers whose productivity has not altered. Esteem is not modelled as a zero-sum game in status. In the present example, since half of the workforce experiences a substantial average productivity increase of 25 per cent, averaged over the entire population the direct gain in esteem is 0.125.

However, this direct gain is offset by indirect losses resulting from the decisions to switch salience. For the critical worker who chooses to switch identity from nationality to job, W = 0.809, whereas N = 1.125. She is therefore getting a gain from the objective job identity of 0.1545, but a loss from salience of 0.316. Were she not to switch salience, she would still get the gain in objective job identity of 0.1545, but suffer no loss from her choice of salience. Hence, for her to be rationally indifferent about the switch, the gain in esteem from membership of the new class rather than remaining in her former class, must equal 0.316.

Where does this difference in class esteem come from? We know if the critical actor does not switch salience, there is no change in the esteem generated by objective national identity, nor from her choice of salience that enables nationality to confer an additional subjective esteem. The objective esteem from job identity is also the same regardless of

salience. The big difference made by remaining with nationality as salient, comes from class esteem. Recall that this depends upon the average esteem among members of the group, of what is generated by their nationality, job, and salience, (each weighted by one third). The objective esteem for the average member of the class from nationality is the same regardless of the choice of class, but objective job esteem is now radically different. Having chosen nationality, the average for the group is 0.2683.[3] If, instead, the critical actor had chosen job, the average for the group is 1.66.[4] Hence the difference in the contribution of objective job identity to class esteem is 1.392/3 = 0.464. This is what is making the difference. This large opportunity cost of persisting with nationality is partially offset by the larger contribution made directly by salience, to bring the net loss to 0.316.

To see the overall effect on wellbeing, we can aggregate four distinct groups of the population. The top 18.25 per cent of wage earners end up with a considerable average gain. Their average earnings, and hence their objective job esteem, rise by 0.41. They have no change in objective esteem from national identity, and they make a direct gain from switching salience of 0.375/2 = 0.19. Their absolute gain from their new class identity is the average of absolute gain for the class of those who switch salience. To work this out, we first need to calculate the effects on the other component of the new class. As we will see, it is 0.11. Summing the four components of esteem, the top group gets a hefty absolute increase in esteem of 0.71.

Now consider the remaining 16.3 percent of the population who switch salience, who switch despite getting more esteem from the nation that their job. Their gain in objective job esteem averages 0.24. Their objective esteem from nationality is unchanged, and they make an average direct loss from switching salience of -0.16. The average change in the esteem from the new class identity is the weighted average of the two classes from each of the three direct sources of esteem. So, the average gain for the class from objective job identity is 0.32; and from the change in salience is a tiny 0.02, with no change in that from nationality. Hence, the absolute change in esteem from class increases by 0.34/3 = 0.11. Again summing the four components, the net gain for this group is 0.19.

The next group is the remaining 15.45 per cent of the population for whom productivity increases but salience is not switched. For them, the objective increase in job esteem averages 0.04. The direct objective and subjective contributions of nationality are unchanged. Prior to the increase in wages for the upper half of the population, they were in the same class as everyone else, and received the reflected glory of average productivity of 0.5. Now, the average productivity of their class has fallen to 0.26, the slight increase from the unchanged average productivity of the bottom half of the population, with whom they have chosen to remain in the same class, being due to the small increase in that of their own productivity. Hence, they get a loss of esteem from the productivity of the class of 0.24, weighted by one third, namely -0.08. Summing the four components of esteem, the absolute change for this group is -0.04.

[3] 0.25 + {0.07725[1 – (15.25/65.25)]}
[4] (1.5 + 0.809)/2

The final group is the remaining 50 per cent of the population for whom nothing changes except the contribution of class identity. In absolute terms only one component of this changes, namely the objective job esteem of the class. Prior to the increase in wages for the upper half of the population, they were in the same class as everyone else, and received the reflected glory of average productivity of 0.5. Now, the average productivity of their class has fallen to 0.26, the slight increase from their own unchanged average productivity being due to the small increase in that of the third group. Hence, they get a loss of esteem from the productivity of the class of 0.24, weighted by one third, namely -0.08. For this bottom group, the absolute change in esteem is simply this last component, -0.08.

Weighting each of these effects by the share of the group in the population, the total increase in the esteem of the population is 0.12. In comparison, were the society to remain united, the absolute gain in esteem would be 0.17. To put this in perspective, the initial level of aggregate esteem, summed over the four components, is 3.75.

Pulling this together, nearly 30 per cent of the potential gains in the total esteem of the population from the rise in productivity have been dissipated *because the most productive third of the population has chosen to withdraw from shared identity*. In doing so, an elite of less than a fifth of the population has captured more than the entire increase in total esteem, gaining almost as much from switching its identity as it does from the direct contribution of the additional pride in its higher productivity. The remaining fourth-fifths of the population in aggregate suffers a small absolute loss in esteem, despite some of its members getting enhanced pride from their own increase in productivity. The switch in the salient identity of the most productive thus substantially enhances their own wellbeing at the expense both of everyone else and of national wellbeing.

Note that far from the assumptions of the model being stacked in favour of finding that an increase in wage inequality inevitably produces a loss of esteem among the less productive, it assumes that even those who are left out of the increase in productivity are willing to get an increase in their own esteem from the reflected pride of association with those who have become more productive. Far from assuming envy, the model assumes a generous disposition to enjoy, vicariously, the success of others. It is the successful who block this by setting themselves apart and denying shared identity.

6. Extensions

I now consider possible extensions of the model, some evident from the model as currently set up, and some taken from the literature discussed in Section 2.

Class esteem as endogenous

In the above model, each of the four components of esteem, nationality, job, salience and class, is given equal weight: while the values of each component have changed, the weights on these values have been constant. Here I revisit the assumption that the weight on class is exogenous. A possible way in which the weight on class might be endogenous is for it to depend upon the difference in esteem between the classes. Arguably, the larger is this

difference in esteem, the more salient does class itself become, as distinct from nation or work. The fundamental equation, (6), would be modified by the addition of a term β:

$$2N + Wr_c + \beta[2N + (Wr_c/2)]/3 = N + 2Wr_c + \beta[N + W + Wr_c)]/3 \qquad (11)$$

β would itself be an increasing function of the difference in group esteem:

$$[N + W + Wr_c] - [2N + (Wr_c/2)]. \qquad (12)$$

The consequence of this extension is straightforward: for any of the exogenous changes considered above, it amplifies the size of the resulting switch in salience. The exogenous changes, such as a reduction in national prestige, increase the esteem gap as set out in (12), and this in turn now increases β, which thereby increases r_c as implied by the change in (11). Hence, the assumption of equal and exogenous weights has likely biased downwards both the efficiency and distributional consequences discussed.

Inter-generational transmission

Now consider how the behavioural and normative implications of incorporating the Bisin-Verdier model of inter-generational transmission into the present set-up. In the present set-up the cultural trait to be transmitted between generations is the choice of salience. In the initial equilibrium, everyone makes the same choice and so all parents will rationally leave transmission to the costless public good of random social interaction. As a result of the increase in wage inequality, the equilibrium changes, with the top third of the population adopting the new trait of making their high-paying job salient. This group then becomes the cultural minority in the Bisin-Verdier model, with the clear prediction that it would start to invest in the two costly channels of cultural transmission: more intensive parental interaction with their children, and greater control of social interaction, reducing child contact with the majority group. The key prediction is that this would not become a general trait across the entire society: the increased effort by the minority would be distinctive, albeit possibly inducing a smaller defensive increase in effort by the majority.

This is a testable proposition and it is fully consistent with the evidence on the change in child-rearing practices between the high-earning educated elite and the rest of the population. In Britain, both groups of parents have increased their hours of interaction with their children since the time when the dispersion of wages was narrower, but the increase has been dramatically larger in educated households (Wolf, 2013, and Sullivan and Gershuny, 2012). Putnam (2016) provides an extensive array of equivalent evidence for America. The new job-salient class is investing far more household resources, in direct transmission than the previous generation of those earning relatively high wages. It is also investing far more in indirect transmission. As Putnam shows, differences in schools are less significant than might be imagined in respect of their function - the acquisition of cognitive knowledge; but more significant as sites for social interaction between children. The key channel for oblique transmission is the purchase of housing within the catchment area of a school. By diverting expenditure into housing, high-class parents have reduced the social interaction of their children with low-class children.

The Bisin-Verdier model provides a ready mapping from this change in behaviour to the normative implications: the reduced reliance upon the public good of random social interaction is socially wasteful. Further, as wage inequality increases, the size of the minority increases. While there were evidently many influences on both the Brexit and Trump votes, they can reasonably be interpreted as crude proxies for the current size of the two cultural identities. Each society has been revealed as being divided down the middle. In the Bisin-Verdier model, this is the peak level of social inefficiency at which both classes are driven into large direct and oblique expenditures on cultural transmission.

Endogenous altruism

In the basic model, all actors might be considered 'weakly altruistic', in the sense of getting some utility from the wellbeing of the average member of the group with which they subjectively identify, while being indifferent to the wellbeing of other members of the society. This is pro-sociality to other members of the group is consistent with the notion of 'belonging'. That the choice of group is rationally based on individual utility maximization is not in tension with this characterization. Those who opt into the job-salient group, still value belonging to that group, and hence are weakly altruistic towards its other members. However, the model could readily be refined to introduce different degrees of pro-sociality for members of the two groups: for example, those in wage-salient group might adopt a package of beliefs that espouse self-fulfilment, and other forms of selfishness, over all forms of care for others.[5] Williams (2017) cites evidence for such a divergence of values in America. The distinctive dimensions of morality regarded as most salient by ordinary workers are 'protecting', 'interpersonal altruism', and sincerity; among the professional class, while none of these is salient, the distinctive addition is 'self-actualization. If the switch of high-earners to job-salience were part of such a wider switch to a new package of beliefs, it would tend to reduce the incentive for those who would directly lose from switching salience, ($N>W.r_i$) [CHECK] because they would value less the offsetting gain of joining the elite group.[6]

A further possibility is that the two salience groups develop oppositional identities in which indifference to the wellbeing of members of the other group degenerates into gaining pleasure from harming them. Hjort (2014) demonstrates a social context in which this appears to have happened. In now turn to the political consequences of changes in identities, where such a change would have clear and adverse implications.

Endogenous politics

Both the Shayo and Besley models incorporate the consequences of changes in identity for politically-set public policies. Whereas in the Shayo model, the attachment of the low-wage class to nationality rather than class results in policy change that reduces income redistribution, in the present model the change in identity comes from the abandonment of shared salient identity – 'nation' – by the elite class. This is the process characterized and

[5] I discuss this concept of belief packages in the context of the defection of elite wage-earners to job-salience in Collier (2018).

[6] I am indebted to a referee for raising this possibility.

documented in Britain by Goodhart (2017) as the rejection of placed based identity by the elite ('everywhere people'). Rueda (2017) and Munoz and Pardos-Prado (2017) analyse the political implications of a rejection of shared identity by the elite. Using different empirical methodologies – survey evidence versus lab experiments in framing choices – they each find that such rejection reduces the willingness of above-median earners to pay taxes for redistribution to below-median earners. On this interpretation, the observed reduction in political support for redistribution is due not to the increase in nationalism among the low-wage class, as argued by Shayo, but to the rejection of shared national identity by elite wage earners, as modelled in this paper. In the Shayo model, the poor majority automatically gets its way in policy-setting. The present paper recognizes the possibility that even in democracies elites may be disproportionately influential. Thus, there are potentially two distinct political routes to redistribution, a class war in which the poor are victorious, as envisaged by Shayo; and an equilibrium of reciprocal altruism dependent upon social cohesion, (as analysed by Rotemberg (1994)), which is secured by the commitment technology of shared salient identity. Shayo interprets the exceptionally equal income distribution of Scandinavia as evidence of the former, but it might equally be interpreted as evidence of the latter. Similarly, the apparent rise of nationalism might be a reaction of those who have retained place-based salient identity, to the exit from shared identity by the elite: their nationalism may not be a deluded retention of shared identity, but its transformation into an oppositional identity focused on immigrants. Existing survey data on identity is problematic: when an identity is universally shared, it is likely to be less prominent in survey responses than when it is has become an opposition identity. In the present state of the empirical evidence, neither interpretation can be decisively rejected and new data, explicitly designed to distinguish between them, is probably necessary.[7]

Generalizing the model

The set-up assumed in this paper has sacrifed generalization for tractability. The advantage has been that the model has been able to generate new welfare effects and identify 'critical actors' in their clearest possible form. But this has come at a cost. While it has been an advance to move from a wage distribution in which there are only two wages, as in the Shayo model, to one that assumes a continuous distribution, the assumption that it is initially uniform and compared with a stylized redistribution, is highly artificial and needs to be generalized to a log-normal distribution with variable parameters. Similarly, it is useful to have a choice as to which identity is made salient, but the assumption that salience doubles utility from the chosen identity needs to be replaced by a variable premium which is endogous to some social process. More general modelling would yield two benefits. It would provide a better sense as to the circumstances in which the adverse redistributions of wellbeing identified in this paper are likely to matter. Moreover, they could then be investigated empirically, integrating conventional economic data on wage distributions with social psychology evidence on identities and their salience.

7. Conclusion

[7] I would like to thank the psephologist Steven Fisher for this point.

It is no longer controversial for an economic model of behaviour to incorporate both choices of identity and the esteem generated by that choice. In this paper, I have combined objectively determined identities of nationality and job, with a choice as to which of them should be made subjectively salient. This choice itself potentially creates a cleavage between those who choose nation and those who choose job, which gives rise to a further difference in esteem: in effect, the model endogenizes the new widely noted class formation between nationalists and the skilled (Chua, 2018, Goodhart, 2017, Williams 2017). What the model brings to the analysis are three results that are arguably non-obvious. First, a small change in the distribution of a continuous variable can produce bifurcation into group identities. Second, such a change has both efficiency and distributional effects. Third, while these effects are generated by privately optimizing behaviour, on a conventional Utilitarian metric both are adverse: there is an overall loss of efficiency, compounded by a regressive redistribution.

References:

Akerlof, G. and R. Kranton, 2011, *Identity Economics*, Princeton University Press.
Akerlof, R., 2017, Value Formation: the Role of Esteem, *Games and Economic Behavior*, 102, 2017, 1-19.
Benabou, R. and J. Tirole, 2011, Identities, Morals and Taboos: Beliefs as Assets, *Quarterly Journal of Economics*, 126, 805-55.
Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic Theory*, 97(2): 298–319.
Chua, A., 2017, *Political Tribes*.
Chandra, Kanchan. 2012. *Constructivist Theories of Ethnic Politics*. Oxford University Press.
Collier, P. 2018, *The Future of Capitalism: Facing the New Anxieties*, Allen Lane.
Collier, P., 2016, The Cultural Foundations of Economic Failure: a Conceptual Toolkit, *Journal of Economic Behavior and Organization*, pp. 5-24.
Eguia, Jon X. 2017. "Discrimination and assimilation at school." *Journal of Public Economics*, 156: 48–58.
Goodhart, D., 2017, *The Road to Somewhere*, Hart.
Hjort, J. 2014, Ethnic Divisions and Production in Firms, *Quarterly Journal of Economics*, 129, 1899-946.
Hood, B., 2014, *The Domesticated Brain*, Pelican.
MacDonald, K. and T. M. MacDonald, 'The peptide that binds: a systematic review of oxytocin and its pro-social effects in humans', *Harvard Review of Psychiatry*, 18.1, 1-21.
Muoz, J. and S. Pardos-Prado, 2017, Immigration and Support for Social Policy: an experimental comparison of universal and means-tested programs, *Political Science Research Methods.*
Putnam, R.D., 2016, *Our Kids: the American dream in crisis*, New York, Simon and Schuster.
Sapolsky, R., 2017, *Behave: the biology of humans at our best and worst*, London, Bodley Head.
Rotemberg, J., 1994, Human Relations in the Workplace, *Journal of Political Economy*, 102(4), 684-717.
Rueda, D. 2017, Food comes first, then morals, redistribution preferences, parochial altruism and immigration in Western Europe, *Journal of Politics*.

Shayo, Moses**,** 2009. “A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution.” *American Political Science Review*, 103(02): 147–174.
Sullivan, O., and J. Gershuny, 2012, Relative Human Capital Resources and Housework: a longitudinal analysis. Sociology Working Paper 2012-04, Department of Sociology, Oxford University.
Williams, J. C., 2017, *The White Working Class: overcoming class cluelessness in America*, Harvard Business Review Press.
Wolf, A., 2013, *The XX Factor: How the rise of working women has created a far less equal world.* New York, Crown.