

The Cultural Foundations of Economic Failure: a Conceptual Toolkit¹

Paul Collier²

1. Introduction

This paper proposes a simple analytic framework by means of which the concept of dysfunctional ‘culture’ can be studied rigorously and in a unified way by social science. A dysfunctional culture delivers outcomes that are the antithesis of ‘thriving’: people are trapped in varieties of misery. Yet these extremes cover the range of behaviour that conventional economics is least-well equipped to analyze.

Dysfunctional cultures are found across the various levels of social aggregation. For example, even in the relatively culture-light setting of American manufacturing firms, organizational cultures have recently been proposed as the primary explanation for persistent large productivity differences (Gibbons and Henderson, 2013). I will focus on three disparate levels of aggregation: countries, firms, and families. Just as Gibbons and Henderson demonstrates that there are persistent differences in performance between firms, and attributes them to culture, so Clark (2014) demonstrates that there are highly persistent differences in performance between families, and attributes them to either culture or genetics. Belatedly, culture is now also being studied as a contributing factor to the persistent differences in performance between countries. *Mind and Society* (World Bank, 2015), and Spolaore and Wacziarg (2013), are valuable analytic surveys of this work. My aim is to provide an economic account of culture which is applicable across these three very different social contexts, while being sufficiently specific to be operational.

I use the word ‘culture’ because it approximates more closely than other terms the concept I wish to explore. It invites ‘balance’ into analysis by accepting that both knowledge and preferences are socially generated and that self-interested rationality is a component of behaviour rather than its exclusive determinant. Only by radically enlarging its account of behaviour can economics hope to provide an adequate account of severe social dysfunction. But I am acutely conscious that the ordinary language meaning of ‘culture’ is much broader than the concept I develop, and it also carries much cognitive and emotional baggage. The concept of culture which I use is precisely defined by its constituent components; namely beliefs and social networks. Behaviour is posited to be generated by decisions that are rational conditional on incentives and beliefs. In turn, beliefs are acquired through participation in social networks which are often pre-rational. The orthodox picture of Bayesian rational self-interested individuals interacting without structures other than those strictly implied by resource constraints needs to be replaced by a picture of individuals as naturally selected organisms living in structured populations, whose behaviour faces cognitive and physiological, as well as environmental, constraints. At the time

¹ I would like to thank George Akerlof, Tim Besley, Colin Camerer, Vince Crawford, Steven Pinker, Nick Rawlins, Paul Seabright and Dennis Snower for comments on a previous draft.

² Blavatnik School of Government, University of Oxford, and International Growth Centre.

when a culture forms it favours the individuals who adopt it, but for the group it may become highly dysfunctional.

The actor is posited to hold three types of belief. One, *identities*, influences preferences: once an actor has adopted an identity, this will shape decisions, perhaps more powerfully so than the effect of incentives. The second, *narratives*, influences how the actor understands the causal relationships by which the consequences of particular actions are determined: narratives may trump the direct observation of experience as means of learning about the world. The third, *norms*, influences the constraints which the actor has internalized: an actor who believes that violence against others is wrong will behave differently from one who thinks that violence is a source of pride.

These three types of belief held by an actor have various sources. Those most conventionally considered in social science are direct observation of the world and genetic inheritance. While these are undoubtedly important sources of beliefs, as explanations of cultural dysfunction they are unpromising. Quirks of evolution can explain some universal decision biases, but not behavioural differences between cultures. In the proposed structure, beliefs are initially generated, at least in part, through participation in *social networks*. This participation is pre-rational, and indeed involuntary. The networks into which an individual is born and raised are therefore treated as primary: evolution, both of beliefs and participation in networks, is path dependent.

The culture that characterizes a context (such as an organization) is locally stable even if dysfunctional. This stability comes from two distinct processes. One is that in equilibrium the different types of belief are mutually reinforcing: for example, norms support identities. The other is that the networks in which an actor participates reinforce his beliefs through various means. The core of the paper elaborates on this structure.

The proposed framework imposes a common testable structure to the analysis of any particular form of cultural dysfunction. First, behaviours are the result both of incentives and beliefs. Second, beliefs are generated not just by genetics and experience, but by participation in networks. Third, in each context, principals - CEOs in respect of organizations, heads of households in respect of families, and governments in respect of societies - may wish to improve these outcomes. The instruments at their disposal for behavioural change are both the conventional array of incentives, and strategies that influence cultures. Principals have the scope to influence a culture both by directly changing particular beliefs, and indirectly by changing participation in networks. One of the frontier issues in social science is how incentives interact with other means of changing behaviour. The framework provides a common and comprehensive analytic checklist for studying this interaction in particular contexts.

The framework is less parsimonious than that of conventional economics in two respects. A relatively minor departure is that norms are not subsumed into preferences, but given a distinct

role. This follows a very long tradition in social thought in which what we desire to do is distinguished from what we believe we ought to do. Economics has conventionally depicted this tension as being played out between a self-interested individual and an altruistic Utilitarian social planner, but it is fundamentally a personal psychological tension that is worth preserving. More radically, behaviour is no longer entirely grounded in self-interested rationality. Rationality only shapes decisions once many beliefs have been uploaded from pre-rational network participation.

Parsimony is a huge advantage, and in some instances self-interested rationality is adequate to explain dysfunctional outcomes. However, in many others the constraint of rationality becomes an encumbrance rather than a valuable discipline. Economics is already willing to countenance modifications. One has been to introduce biases resulting from the neurology of decision-taking (Kahneman). Another has been to broaden the concept of self-interest to include concern for how others regard our behaviour (Benabou and Tirole). A third, closest to the present approach, has been to root behaviour in internalized identities and norms (Akerlof and Kranton). The proposed framework can be thought of as an extension of this latter approach, adding a further belief alongside those of identities and norms, namely (mis)understandings of causality acquired through narratives, and grounding all three of these beliefs in pre-rational networks. While the framework is less parsimonious than conventional economics it remains highly reductionist and far from being a definitive depiction of social behaviour. Rather, it is intended as a staging-post in the evolution to a new trade-off between parsimony and realism.

The rest of the paper is organized as follows. Section 2 sets out in greater detail each component. The three types of belief are inter-dependent with each other and with networks, but these interdependences are temporarily set aside. Section 3 introduces the interdependences. By definition, the beliefs generated by a dysfunctional culture deliver outcomes which are inferior to those which could be generated by alternative beliefs. The interdependences explain why, despite this continuous stream of observable evidence, such cultures may nevertheless be highly stable. Section 4 applies this structure to three contexts in which cultural dysfunction matters: societies, organizations and families, and discusses how principals can use culture to change outcomes. Section 5 concludes with some potential applications and a suggestion for how cross-fertilization across research which tackles the same mapping in different contexts might be realized.

2. Beliefs and Networks

2.1 Identities

I begin with the belief in identity since this has been the subject of much recent economic theory. The least disruptive approach is to enter identity directly into the utility function alongside consumption and work. In their characterization, identity is a private asset that enters into utility in the same way as a consumer durable. This successfully captures why an actor might rationally choose to protect the value of the asset by screening out potentially disruptive information. An

important corollary is that an identity is a stock accumulated from prior flows. However, in two respects it is inadequate.

First, a shared identity can reasonably be seen as a fundamental human trait. Normal human beings need a sense belonging to a group to give meaning and coherence to their lives. The lack of such a sense is a mental illness: Asperger's Syndrome and autism. Identity is a constituent component of what it means to *be* human rather than being a consumer durable. Identity is not a commodity which is part of our preference orderings, but rather a psychological disposition which generates our preference orderings. It is a characteristic the value of which is incommensurate with the domain of economic choices. Different identities generate different behaviours. This is the key point of Akerlof and Kranton (2011) who treat identity as a preference-changer. Much psychological evidence supports this interpretation. Economics has long treated preferences as a constituent component of being human. Now that economics has to come to terms with the importance of identities, rather than see them as a previously neglected consumer durable, it would be more fruitful to see them as the DNA of preferences.

Second, for an identity to be meaningful it must claim membership of some bounded group. The actor is neither unique nor undifferentiated, but belongs to a proper subset of humanity. In this respect, an identity is a club good. The club for which an identity is a good is a social network of people who share the same identity and in which the actor participates. One function of a club may be that participation is *performative*: expressing an identity, an example being the expressive theory of voting. Further, while voting is exceptional in being secret, most performative actions in networks are observable, and so generate information which circulates around the network. As this information is received by other participants it reinforces that identity.

The importance of identity as an influence on behaviour is well-established. It is even apparent in infancy: a four year old primed to think of himself as 'a guard' will stand still for four times longer than one primed with less pertinent identities when faced with the same incentive (Seligman, 2011). The most heavily researched aspect of cultural transmission of behaviour has been the effect of exposure to violence on television. Here the evidence of adverse effects is overwhelming (Dijksterhuis, 2005). Akerlof and Kranton, (2011), show that a worker who adopts the identity of an insider will perform better than one who adopts the identity of an outsider when faced by common economic incentives. Their insider/outsider distinction applies more generally than at the level of employment. A teenager who has adopted an outsider identity of revolt will be less supportive of other family members than one who adopts the insider identity of a member of a multi-generational family. Similarly, a citizen who adopts the outsider identity of a foreigner will be less inclined to cooperate with other citizens than one who adopts the insider identity of a national. For example, using the methodology of a controlled experiment, Candelo-Londono *et al*, (2011) find that Hispanic immigrants to America who have learnt English are more willing to contribute to public goods. Once adopted, identities may largely shape behaviour: economic incentives may have relatively little effect. The young adult who has

adopted the identity of a drug dealer will be maximizing subject to incentives over a radically different range of considered choices from the one who has adopted the identity of a Quaker.

The key issue is not whether identities are important for behaviour, but how they are acquired. The adoption of an identity is not part of the general process of selection of commodities subject to preferences and prices.³ Imitation appears to be a powerful psychological process, although quite how powerful it is remains open to debate. The most straightforward mechanism, which is hard-wired into the human brain, is the copying of an observed action. Flowing from the discovery of the mirror neuron, we now know that the perception of an action and its performance are neurologically related. But copying observed actions is merely the ‘low road’ of imitation; the high road is imitation of stereotypes. The evidence here is striking, though contested (Dijksterhuis, 2005; Hood, 2014, p.97). Teenagers primed to think of the attributes of professors perform significantly better in tests than a control group primed to think of football hooligans. People primed to think of rudeness behave significantly more rudely than those primed to think of politeness. People primed to think of the elderly walk more slowly. Asian-American women perform better on maths tests when their Asian, as opposed to their female, identity is emphasized.

As an approximation, identity is a characteristic which is acquired pre-rationally through the imitation of stereotypes resulting from participation in networks. Intervening in the process by which identities are adopted may offer more scope for a principal to achieve an objective than the manipulation of material incentives. Although the typical person participates in several networks and so has multiple identities (Sen, 2006), few identities are accessible to everyone. Many will only be accessible in networks in which the person does not participate, and others will be incompatible with an identity that has already been acquired.⁴ Family networks differ in the functionality of the stereotypes with which they endow their members. Hence, the stereotypes accessible through the relatively few open-access networks are a form of public capital of most potential benefit to those whose private network endowment is least functional.

2.2. Narratives

The causal theories that people believe, determine what they perceive to be maximizing behaviour given their identity-based preferences. People acquire understanding of causal processes partly through exposure to analytic explanations, partly through exposure to narratives, and partly through the direct observation of events. Directly, narratives are stories of actual events that merely amplify direct observation. Potentially, this merely accelerates a Bayesian

³ Sen (2006) claims the opposite: several times in the book he invokes a direct analogy between choices about identity and choices over budgets. But the informational requirements for informed choices of identity are an order of magnitude more demanding than for choices between commodities. This is why, as Kahneman (2011) argues, copying the behaviour of others may be a rational strategy. By the same reasoning, the agent has little basis for selecting which role model to copy beyond relative exposure.

⁴ On the edges of identity, people can find themselves with more than one identity in the same frame of reference, but we think of these situations as posing a dilemma: hence the concept of a ‘conflict of interests’.

process by which events cumulate into causal hypotheses. However, narratives may predominantly be useful hooks on which to hang the generalizations necessary for prediction and intervention. Evidence from psychology shows that the mind retains narrative-derived understanding better than that derived from analytic understanding: narratives trump analysis. Zak (2014) suggests that those narratives which involve a struggle by an individual to whom the hearer can relate, which are expressed lyrically, and which are repeated, are the most memorable. Particular events may become iconic narratives, functioning as generic formulae, even if they do not encapsulate the most reliable statistical relationship. They may simply be more memorable, or they may be circulated in networks disproportionately either through chance, or because nodal actors use them strategically to influence behaviour.

Because narratives are delivered through networks whereas direct experience is individual, narratives are reinforced by sociability. In discussing networks I note evidence that people are inclined to trust network-delivered narratives above the evidence of their own eyes. Hence, narratives appear to be the predominant vehicles for understanding, (or misunderstanding), causal processes.

At the level of the individual, the concept of narratives is standard in psychology. Wilson (2011) discusses extensively how dysfunctional adolescent behaviour can be ‘redirected’, through the drip-feed of disruptive narratives. At the level of society it is a key concept in media studies: narratives are how communications professionals in politics and business seek to influence behaviours such as voting and consuming. In economics the idea goes back at least to Keynes’ theory of probability and is applied to financial markets by Akerlof and Shiller (2009). As they put it, ‘stories no longer merely *explain* the facts, they *are* the facts.’⁵ Mass market novels can both encapsulate and change the narratives believed by a society. Zingales et al., (2006) use the evidence from the narratives of novels that are popular in different societies to demonstrate that narratives vary considerably between societies.

The causal hypotheses encapsulated in a narrative need not be correct in order for it to be believed. Mullainathan *et al.* (2006) discuss a variety of false theories which result from what they term ‘coarse thinking’. Causal processes may often be understood through categories: for example, ‘America wins its conflicts’. Coarse thinking arises because people group situations into categories and sometimes mis-assign them. Such mistakes can sometimes be the result of deliberately misleading persuasion. They give the example of a political speech in favour of free trade by Arnold Schwarzenegger in which he depicted trade as a struggle and argued ‘Now they say India and China are overtaking us. Don’t you believe it. We may hit a few bumps – but America always moves ahead. That’s what Americans do.’ False theories will gradually be superseded, but the process of error correction can be very slow. They may be actively promoted by those they advantage; counterfactuals are often not observable; some adverse evidence is filtered out by taboos; and if permitted evidence is ambiguous it is reweighed so as to support

⁵ Akerlof and Shiller (2009), p54.

existing beliefs (Haidt, 2012). Hence, people will be maximizing subject to beliefs about causal relationships which are inaccurate to varying degrees.

Some narratives have direct consequences for behaviour. An example of a generic narrative that directly affects voting choices is 'government wastes money'. Since narratives can diverge from reality, they are liable to differ between societies. Thus, a society in which voters believe this narrative will have lower levels of public spending and taxation than one which does not. The narrative of fatalism -'it's not worth trying'- is a loss-aversion (disappointment-reducing) psychological strategy. A society in which this narrative is prevalent, for example because its members have experienced failure and so find loss-aversion particularly attractive, will thereby be characterized by lower effort than a society characterized by a narrative of opportunity. One reason why Africa had failed to attract more private finance for infrastructure is that the prevalent narratives discourage making deals with foreign companies. Transactions are commonly seen as being zero sum games, and a legacy of colonialism is that foreigners are widely seen as exploitative.

2.3 Norms

To the extent that economics has incorporated norms, they have usually been considered as acquired preferences. Postlewaite (2010) provides a recent survey of the economic perspective on social norms. In his framework they are categorized as preferences and divided into two types, being either primary or consequential upon some feature of society such as a missing market. The two canonical examples used by Postlewaite are that he would not steal his colleague's pen, even if the theft would not be detected, and that tastes as between baseball and reading the Talmud differ. But these examples illustrate a manifest difference. The first is indeed a norm, but it is not really a preference. Rather, it is an internalized constraint upon choices: he wants his colleague's pen, but this want is frustrated by a moral impediment. The second supposed 'norm' is a straightforward difference in preferences: it cannot sensibly be viewed as a difference in constraints on choice. However, as Postlewaite acknowledges (and perhaps why he misclassifies it as a norm), it is a difference in preferences that is generated by differences in internalized identities. I think it is both clearer and more in harmony with the conventional meaning of language, to preserve the distinction between preferences and norms while endogenizing them in deeper social processes. Thus, with preferences set by identities, I suggest that norms be conceptualized as socially constructed constraints on behaviour. In societies with the same structure of incentives and preferences, but different norms, people will make different choices.

Besley, Jensen and Persson (2014) provide a simple yet fertile account of precisely how norms influence behaviour and apply it to tax evasion. In their model there is a tension between preference and norm: people prefer not to pay tax but believe that they should do so. Individuals differ in the intensity of their intrinsic moral beliefs. Each individual participates in one of

several social networks, between which there are also differences in the mean intensity of moral belief. How this tension between preference and norm is resolved is determined by two consequences of the variation in moral intensity among the population. The most evident is that the stronger is the individual intensity of moral belief the more likely is the actor to prioritize it over his preference. This is the emotion of guilt. However, there are two further effects which can be critical: the actor does not want to adopt behaviour that is bad relative to that of his network. One reason to avoid such behaviour is stigma, which is determined by the reactions of others in the network. Between guilt and stigma there is an intermediate emotion of shame: for even if stigma is avoided because others do not observe the behaviour, the actor cannot avoid revealing to himself a negative *self*-evaluation. Were the actor to prioritize preference over norm, it would differentiate him from others in the social network to a degree determined by the proportion of other actors who reveal the same behaviour. On minimal assumptions, the greater is the proportion of people who breach the norm, the narrower is the difference in the intensity of moral belief between those who breach it and those who do not and this generates clear dynamics. While for an issue such as tax compliance the relevant portion of the distribution of behaviour is usually where non-compliers are a minority, for other behaviours such as charity the other part of the distribution may be pertinent, with its corresponding emotions of virtue, pride and praise. The intensity of the moral beliefs is assumed to differ over the population. It is also assumed to change in response to external events: the introduction and subsequent withdrawal of the poll tax temporarily shifts the mean of moral beliefs in some networks but not others.

Robert Akerlof (2013) also builds on the idea that people adhere to norms partly because they have internalized the norm and partly because others in their social network have done so. He argues convincingly that norms are grounded in a sense of what the 'reasonable person' should do in a situation. Whereas Besley *et al.* use the social network effect to generate the psychological power of shame, Akerlof uses it to generate anger. If a reasonable person should follow the norm, then anyone who breaches it provokes anger in observers: that anger motivates punishment which is what generates stigma. In turn, the expectation of stigma reinforces adherence to the norm even by those who have not internalized it. This is self-reinforcing, because if almost everyone is adhering to the norm, then a 'reasonable person' should indeed do so.

This self-reinforcing nature of norms implies that while getting a new norm established might be difficult, once a norm is established it is liable to be a locally stable equilibrium. Many norms are arbitrary: 'drive on the left' and 'drive on the right' are incompatible norms with no intrinsic moral ranking. If, like 'drive on the left' and 'drive on the right', many incompatible norms are locally stable, we should expect the norms that are adopted by a society to be path-dependent. Random variations merely in the sequence of the same events between societies would give rise to permanent differences in norms. Sensitivity to sequence is one of the key results of the analysis of networks. In effect, this is what the analysis of Besley *et al.* demonstrates. The

content of the norm is immaterial to the analysis which depends upon exploiting properties that follow from random perturbations to the distribution of adherence across the population.

However, many norms are not arbitrary. They may follow fairly directly from a limited number of moral values as suggested by Haidt (2012), who has recently revived analysis of fundamental values of ‘the righteous mind’. He postulates that six values are common to virtually all cultures, with the typical person placing some weight on each of them: care, fairness, liberty, loyalty, authority and sanctity.⁶ We now know that these moral values are to an extent ‘hardwired’ into human behaviour: natural selection has induced us, in some situations, to generate chemicals that stimulate moral choices (Pagel, 2012; Zak, 2012).

A sense of fairness may help in overcoming the free-rider problem which is the fundamental obstacle in public goods provision. However, other moral values can undermine rather than reinforce collective action. For example, loyalty to group may override the anger that would otherwise be provoked if a member of the same group breaches a norm. Two values collide, and rather than punishing the offender a fellow-member of a group may choose to defend him from non-members who propose to punish. Smith (2005) shows how such behaviour fundamentally weakened state capacity in post-Roman Europe, as loyalty to clan undermined enforcement of the law. For example, ‘the English King Aethelstan was only one of several rulers vexed by criminals whose relatives protected them from the rigours of royal justice. He confronted the problem of thieves who were being sheltered by kindreds so strong and so large that they were beyond the reach of royal officials.’ (Smith, 2005, p.98). This description of 10th Century England applies to parts of 21st Century Africa.

While the analysis of Besley *et al.* is content-neutral, and that of Haidt is value-prescriptive, here I suggest an analysis which is intermediate: the content of norms can be analyzed for their internal consistency with other beliefs. Norms in all societies may have some family resemblance coming from deep values, but they can have persistent differences not only because norms are directly path-dependent, but because they are sustained by belief systems which are path-dependent. An implication is that behaviours should be substantially different when actors subject to different norms are confronted by the same objective incentives. While it is usually difficult to hold all objective incentives constant, Miguel and Fisman (2007) provide just such a test. Their context is parking fines for diplomats in New York which were for some years legally unenforceable, the payment therefore depending upon the internalized norms of diplomats from different cultures. They found that the willingness to pay the fines varied very substantially and were systematically related to variations between national cultures in adherence to the rule of law. There is equivalent research on large differences in trust.

Although people are moral actors, they are also self-serving. They have three potential alibis for breaking a moral norm. One alibi, as in 10th Century England, arises if different values imply

⁶ Pinker (2011) slightly repackages these values into five groups.

opposing choices. The actor can choose to attach more weight to that value which best suits his own interest. A second alibi is if the evidence as to the consequences of an action is noisy. Haidt shows that people weight noisy evidence in a self-serving way. The third alibi is the non-compliance of others. This is an important constraint on the establishment of new moral norms, even if they are readily derivable from common values.

2.4 Networks

Networks are structures of social interaction which channel information. They are often defined as being informal and non-hierarchical, but here I will use the term to include formal, hierarchical organizations, such as firms: while not *merely* networks, they can function as networks. Some of the information channelled through networks is directly about the actual behaviour of other actors: networks are windows on the behaviour of other members of network. Some of the information is about the external world, transmitted between actors.

The most potent networks for the imitation of behaviour may be those which are voluntary but hierarchical. For example, the hierarchy of celebrity on a teenage website may generate deference more potent than relationships of organized authority such as a school. In turn, the structures of deference generate the directions of imitation of behaviour. Voluntarism may be important because participation thereby avoids triggering *reactance*, the psychological phenomenon in which, faced with an order, actors attempt to re-establish autonomy by non-compliance. The type of behaviour observed is usually specific to the network. Being focused on a particular activity, it is potentially well-suited to induce imitation of that particular aspect of behaviour among its members.

As channels of information about the external world networks are powerful. Their only competitor is direct observation and their relative efficacy has been assessed through psychological research. Experiments have shown that even for mundane information received by direct observation, such as the relative size of objects which are in full view, if the evidence of the senses is contradicted by network-sourced information people are more likely to rely upon the network than their own senses. As I now briefly discuss, identities, narratives and norms are not directly observable: *all* information about them is necessarily received through networks.

Identity is by definition about self-perceived relationships to others. Networks are therefore fundamental to identity, not only in being the channels through which information about identities flows, but through themselves being relational. Many different types of information flow through networks and some will not be pertinent for identity formation. But participation in a network is a direct source of identity (Hood, 2014, p.242). Some of the activity of many networks is *performative*. Rituals are the class of activities which are purely performative. Having no other instrumental purpose, they are particularly powerful means of reinforcing an identity because both the actor herself and observers in the network cannot misinterpret it. The

performance of ritual can therefore be understood as an efficient investment in identity, generating virtue in the actor and praise from observers (Hood, p.90).

The information content of a narrative is a readily intelligible theory of some aspect of how the world works. A network is necessarily the vehicle by which a narrative is accessed because the key feature of a narrative is that it is *narrated*. That is, it is designed to be transmitted from one actor to others. Thus, by definition, a narrative is an incident or theory transmitted through a network embedded in a story. People can invent theories, but for theories to become narratives they must be embedded in stories transmitted to others through networks.

Norms, by definition, described how other people behave and expect the actor to behave. Thus they can only be acquired by observing the behaviour of others, or being taught about that behaviour. Both observation and instruction are transmitted through networks.

Omerod (2012) provides a useful discussion of network effects and argues that their consequences need to be better taken into account in public policy. He shows how variations in the structure of networks can have powerful, though not always predictable, consequences for who accesses what information. Two aspects of network structures may be particularly important for our purposes: segmentation and nodal actors.

Segmentation

A few networks reach all agents, and indeed this may be a key reason why they hold their audience: some information has enhanced value if the recipient knows that it is being received by others and so is *common knowledge*. At the level of society, the networks with this characteristic deliver popular culture: that is, the menus of identities, narratives and norms available to everyone.

Such comprehensive networks compete against those with more limited membership. In recent decades the cost of transmitting information has fallen massively. The consequence which is usually emphasised is that people have become better informed. However, another consequence which may be of greater significance is that information networks have become more segmented. Falling fixed costs have enabled networks to proliferate relative to the ability to absorb information. As information networks have proliferated, they have also differentiated, being subject to the same market pressures that have produced product differentiation. The capacity to participate in networks is constrained by time and human capacities, analogous to the Dunbar constant that posits a fixed limit to the number of substantive direct inter-personal interactions. Actors therefore necessarily have to be selective as to which networks they join. The proliferation and differentiation of networks therefore inadvertently generates segmentation and inequality: actors have less common knowledge. British television is a dramatic illustration of this process. Until 1957 there was only a single channel which in consequence everyone watched. The number of channels has progressively increased, while total the audience for television has diminished due to increased competition with other media.

Through participating in different networks, actors hear different narratives and so form different beliefs about causal relationships. Since norms depend upon a common view as to what a 'reasonable person' would do in a particular situation, continued exposure to different narratives will also lead to different norms. Over time, these differences in norms and narratives may lead to differences in identity.

Nodal actors

Networks can have different structures. However, a common feature is that some actors are nodal, in that they are disproportionately connected to many others within the network. Changes in technology have tended to concentrate information around fewer nodal actors, an example being the increasing domination of bestsellers in the book market. Through the power of imitation, the identities and norms of these nodal actors are consequently disproportionately influential as role models in setting the identities and norms of other actors. They are also the most potent disseminators of narratives. The two functions – role models and narrators – can have different weights depending on the status of the nodal actor in the network. For example, since the behaviour of celebrities is celebrated, they are necessarily influential as role models. But whether they are also conduits for narratives, as when film stars promote causes, is contingent. Conversely, since the function of teachers is to transmit information, they are necessarily influential as narrators, but may well not become role models.

Principals who seek to change behaviour through changing narratives need to reach nodal actors (and nodal actors may themselves choose to behave as principals, using their power to alter behaviour in line with their own preferences). Additional to their evident influence as role models, their disproportionate influence over narratives gives them three distinct means of affecting behaviour. First, they may be able to create myths, determining which narrated incidents become elevated into the status of theories. Such potent narratives are analogous in their effects on how causality is understood to the effects of role models on preferences. Just as nodal actors can enhance the potency of an incident, they can reduce its potency. They may be able to *filter out* potentially disruptive incidents by ignoring them. Those incidents that are too prominent to be ignored may be *neutralized* through the generation of propositions that qualify or cast doubt upon them.

Network Participation

To an extent, network participation is a choice. People can choose not to interact with others, but cannot usually require others to react to them: hence, as an approximation, inclusion, which is the option to participate, is not a choice variable, but the actor has freedom to decide whether to exercise the option.

People are born and reared into networks, a few of which are common across a society while many are differentiated and segmented. While young, people will thus acquire somewhat differentiated identities, narratives and norms. As young people acquire a degree of autonomy,

they acquire some freedom to choose their participation in networks. However, the different initial endowments of networks constrain the extent of choice. Even with a modest degree of autonomy, people can seldom be forced to participate in networks to which they would rather not belong. But while a few networks are open access, most can only be joined from a limited number of other networks. However, many of these constraints upon network access may be non-binding. Participants in those networks which build a differentiated identity will generally not want to leave them. A corollary of the theory of taboos developed by Benabou and Tirole (2011), is that just as people screen out information sources likely to be disruptive of their chosen identity, they favour sources likely to be supportive of it. Differentiated identities will thus feed back onto the segmentation of networks. For example, identity shapes which newspapers and television channels people select. The segmentation of networks is therefore liable to generate the segmentation of society.

3. Equilibrium and Equilibrating Mechanisms

The concept of a culture is only useful if cultures have persistence. Although cultures clearly evolve, except in circumstances of large exogenous shocks they do so slowly. Culture provides the missing timescale between that of evolutionary psychology, which is several hundred thousand years, and that of consciously rational self-interest, which rests on calculations over the horizon of the actor's lifetime. Just as physical evolution has converted many aspects of behaviour from consciously rational to instinctive, cultural evolution has introduced pre-rational influences onto aspects of behaviour. Cultures are not hardwired: they are transmitted by social networks, not genetically. But, once transmitted, a culture nevertheless *predisposes* an actor to a particular behaviour. As a first approximation it is therefore useful to think of a culture as being a locally stable equilibrium. For it to be locally stable there must be forces maintaining stability. In this section I explore what these forces might be.

Given the above structure, in which networks generate beliefs, which in turn generate the behaviour which directly determines outcomes, there are a limited number of possibilities. First, the outcomes might feed back upon the beliefs. Second, the outcomes might feed back upon the choice of networks. Third, the beliefs might interact with each other. Finally, the beliefs might interact with the choice of networks.

It is now possible to make more precise sense of the notion that cultures can be dysfunctional. *Essentially, a culture is dysfunctional if the first two feedback mechanisms are too weak to counter the stabilizing net effects of the other forces.* Given this condition, outcomes can be dysfunctional without disrupting the identities, narratives, norms and networks which generate the dysfunction.

3.1 Interactions among Beliefs

Beliefs describe the set of identities, narratives and norms held by an individual. So defined, at the level of the thought processes of an individual, the different components of beliefs interact.

For analytic simplicity it is useful to reduce this to dyadic relationships between the three different types of belief.

Identity – Narrative

Some narratives support some identities whereas others are undermining. The explanation of taboos by Benabou and Tirole (2011) depends critically upon the idea that some narratives can threaten some identities. Psychological evidence finds that people want to identify with success. When asked about a football result, a supporter will typically respond ‘*we* won’ if the team was successful, but by ‘*they* lost’ if the team was unsuccessful (Cialdini, 2007). Evidently, the former narrative reinforces an identity as a supporter of the team whereas the latter weakens it. Narratives of grievance are an important sub-class of influences on identity because they promote identities which are oppositional. For example, the Marxist narrative of ‘the exploitation of the proletariat’ supports an oppositional class-based identity, and nationalist narratives such as ‘the Scots have been done down by the English’ support an oppositional nationality-based identity.

The interaction between narratives and identities can also run in the other direction: an identity can support or undermine a narrative. For example, an identity as a ‘lone wolf’ supports a narrative ‘it’s every man for himself,’ but sits uncomfortably with ‘many hands make light work’.

Identity-Norm

The same forces also apply in interactions between identities and norms. The Akerlof-Kranton analysis of productivity in firms rests on an interaction between identity and norm: ‘I am a good plumber’ matches with ‘a good plumber knows how to do his job’. The identity of ‘I am a gentleman’ implies but does not entail a norm such as ‘I honour my debts’. Reciprocally, someone who holds that norm will find it easier to adopt the identity ‘I am a gentleman’.

Narrative- Norm

Some narratives affect norms. An important class of such narratives is those that transmit a message of grievance or guilt. A sense of grievance, interacted with the bias towards interpreting evidence in the light most favourable to the actor (Haidt, 2012), provides an alibi for the suspension of the obligation to follow norms of reasonable conduct. Conversely, a sense of guilt can overrule the obligation to enforce norms through the punishment of offenders. The narrative ‘A has wronged B’ generates matching grievance and guilt. Regardless of whether the narrative is true, if A and B both believe it, B will have an alibi for breaking norms and A will be disinclined to punish these breaches, further weakening the incentive for B to abide by the norm. Similarly, the narrative that ‘interactions with others are intrinsically zero-sum’, and the norm that ‘people should work together’ are mutually incoherent.

Equilibrium Beliefs

In summary, some dyads are mutually supportive, whereas others are antagonistic. These interactions between beliefs imply that antagonistic dyads are disequilibria whereas supporting dyads are at least locally stable. Hence, observed beliefs will usually be locally stable clusters of identities, narratives and norms all mutually reinforcing. Unstable belief sets adjust: for example, it may be that having adopted an identity, there is a process of selection among possible narratives, so that narratives rather than identities do the adjusting. Or, it might be that norms are primary, for example being learnt in childhood or even hardwired as suggested by Pagel (2012) and Greene (2013), so that subsequent identities adjust to them.

Interactions between beliefs are not the only forces generating local stability. I now turn to a second set of forces.

3.2 Interactions between Beliefs and Networks

Beliefs and networks interact. In Section 2 I suggested that beliefs are generated by participation in networks. I now expand and complicate this assertion. The beliefs that people hold are stocks of ideas which have cumulated from past participation in networks. However, the networks are themselves to an extent chosen as a result of the beliefs that people hold.

In the typical network some participants are more influential than others. At the extremes, in one type of network all participants interact with all other participants with equal influence. In the other, one nodal actor interacts with all other members, but they do not interact with each other.

How networks influence beliefs

Networks are theatres in which people participate in the dual role of actors and audience. As actors they perform a role, as audience they observe the performances of other actors. Both performance and observation are activities that feed back upon the beliefs of the participant.

Networks, like individuals, can be characterized by beliefs. Participation in a network generates several forces that reinforce the beliefs of each participant. The unintended feedback from the behaviour of one actor in a network onto the behaviour of others can be thought of as a special kind of externality, which might be termed a *decision externality*.

To analyse how networks affect norms I take a canonical example of the collective action problem: cows overgrazing the commons. The villagers have agreed that all their cows should be taken off the commons to let it recuperate, but start with nobody adhering to the desired collective action. The timing of each person's journey to the commons to remove their cows is left to individuals and so will be sequential. The dynamics of the material incentives for free-riding in this sequential game are perverse: the more cows that have already been removed by others, the greater the benefit from leaving cows to continue grazing.

Just as the agent who free rides does not take into account the effect of his decision on the *wellbeing* of others, so he does not take into account its effect on the *decisions* of others. The economic cost of compliance increases as the sequence of removal proceeds, but as people remove their cows from the commons, the moral pressure to comply increases. The last person with a cow on the commons indeed has the strongest economic incentive to leave it there, but will be subject to the strongest moral pressure, both communal and internal, to remove it. This psychological pressure is rooted in guilt, shame and stigma. The moral decision externality changes the moral incentive for the actor to leave his cows on the commons because it changes his and his community's judgment of whether the decision is unreasonable. However, while both effects cumulate, the change in the economic incentive is likely to be smaller than the change in the moral incentive. Although the economic incentive to leave cows on the commons is at its maximum in the neighbourhood of full compliance, it remains large even in the neighbourhood of zero compliance. In contrast, whereas in the neighbourhood of zero compliance, not to free-ride would be quixotic, in the neighbourhood of full compliance to free-ride would be psychopathic. That the moral decision externalities cumulate so strongly has important implications. In the neighbourhood of full compliance, collective action may not face a free-rider problem: the moral pressure to comply may be so strong that it overwhelms the economic incentive to free-ride. In contrast, in the neighbourhood of zero-compliance there has been no accumulation of moral decision externalities and so the moral incentive to act is virtually non-existent. Since no other actor has yet behaved ethically there is no moral pressure to do otherwise. To summarize, in a sequential collective action problem each decision generates two types of decision externalities, one on the economic incentive to free-ride and the other on the moral incentive.

The process by which a moral action spreads across a population and so becomes a norm, is analogous to wildfire protests analysed by Kuran (1989). In that model, because there is safety in numbers, each additional person who joins a demonstration reduces the risks faced by other protestors. As a result, the sequential decisions to join a protest can roll across a population. The most motivated join the protest first, and each marginal protester sufficiently lowers the costs to the next potential protester as to switch their decision from non-participation to participation. As Kuran shows, an essential feature for protest to spread is that there should be no wide preference gaps in the population: the more motivated should not be readily differentiated from other actors by some other observable feature. A particular behaviour becomes a norm because the costs of inaction rise: moral decision externalities cumulate.

Because moral pressure is at its weakest in the initial stages of behavioural change, a necessary condition for collective behaviour to be ignited is that some people are willing to be pioneers. They might do so either because they are persuaded that it is morally necessary even if others do not follow them, or because they relish the role. As in the Kuran model, whether the norm then spreads across society depends upon the degree of social cohesion.

Shame and stigma are closely related to *conformity*. People prefer to express opinions which conform to those held by the people with whom they interact. In aggregate, conformity bias increases the proportion of observable behaviour that is consistent with the beliefs of the majority of the network.⁷ Conformity bias applies most obviously in respect of norms. Indeed, norms are predominantly whatever moral codes are normal in the interactions encountered by the actor. An equal participation network of people holding common norms will generate behaviour which reinforces these norms through a quasi-Bayesian process of information updating. Another related psychological process is *imitation*. Imitation may be most important for the reinforcement of identity: the participant imitates the many small markers of identity observed in other members. A third psychological process is *expression*. For example, voting studies find that voting behaviour is often better explained as an expression of identity than by rational self-interest (Hamlin and Jennings, 2011). Some networks provide an arena in which behaviour can be expressive, for example being a football supporter. Allegiance to a team can be maintained in social isolation but is reasonably seen as being reinforced by expressive behaviour. By definition, expressive behaviour depends upon participation in networks of common identity. A participant in a football supporter network reinforces identity both as an actor expressing that identity, and as an observer wanting to conform to and imitate the behaviour of other participants.

The core activity in many networks is to circulate narratives. Narratives are importantly different from identities and norms. The latter are entirely generated psychologically: they have no necessary relationship to objective reality. In contrast, a narrative conveys some aspect of objective reality, albeit potentially in a highly distorted form. A key distinction is between those narratives generated within the network and those generated externally which are then circulated around the network: most networks do both. Internally generated narratives are likely to be compatible with the existing beliefs of the network. Such internally generated narratives are essentially accounts of the behaviour of network members and so liable either to be consistent with network beliefs, or used as vehicles to demonstrate the folly of behaviour inconsistent with network beliefs. The circulation of such compatible narratives reinforces beliefs through the same process as that described above.

Networks as Filters

In contrast, externally generated narratives encapsulate news over which the network has no control. Hence, externally generated narratives are potentially disruptive information. As noted in Section 2, Benabou and Tirole (2011) have recently proposed a theory of taboos in which an investment in identity is protected by filtering out disruptive information. I suggest that the key mechanism by which this filtering process operates is not the individual but the network. Networks filter incoming narratives to varying degrees so as to reduce the disruptive force of

⁷ If behaviour is observed sequentially, paradoxes can arise in which the observed behaviour of the network is determined by the first movers rather than by the beliefs of the majority, which remain unobserved.

potentially disruptive information. This is most readily implemented in nodal networks: here the nodal actor takes on this function. An example is the filtering done by politically aligned newspapers. Typically, aligned newspapers with both under-report disruptive information and propose an interpretation which dilutes its impact. A narrative that has such an effect might be thought of as a *neutralizing proposition*. The nodal actors who generate such propositions are potentially critical to the stability of dysfunctional cultures. In equal participant networks the process of filtering has to be decentralized to the members, but can still be implemented supported by appropriate neutralizing propositions such as ‘information from that source is unreliable’.

The inter-dependence between narratives and norms, and the filtering role of networks can help to elaborate upon the poll tax analysis of Besley, Jensen and Persson (2014). They find that when the poll tax was introduced, tax evasion increased massively in some council districts but not in others, and infer that the norm of tax compliance weakened in those districts. The tax was a classic event which could support rival norms. It could be viewed as unfair because it delinked taxation from the ability to pay – a ‘poll tax’, or as fair because everyone who lived in a community would pay the same amount – a ‘community charge’. An implication of the work of Haidt (2012) is that the choice of interpretation would reflect deeper aspects of personal values, and identities. Accordingly, different networks would filter out one or other of these interpretations. In some districts the new event would be absorbed as being consistent with the norm ‘people should pay their tax’. The tax would be interpreted as fair, analogous to the reinterpretation of a football result by supporters so as to be compatible with identity: ‘we lost because the referee was biased’. In others districts the event would be absorbed as inconsistent with the norm. People would find themselves holding incompatible beliefs and so the intensity of the belief in duty to pay taxes would be undermined as antagonistic beliefs were reconciled.

Potentially disruptive membership

Now consider a person who joins a network with which his belief set is incompatible. There are many reasons why a person might find themselves a new member of a network. Its core function may override any other considerations: for example, the person is a new employee who becomes a member of a network of fellow employees.

If the network is an equal participant network then a potentially disruptive narrative is injected into it. Further, the new participant is unlikely to internalize the norms and narratives that support filtering of potentially disruptive external narratives and so the filtering process is weakened. The new member will also display an identity which is at odds with that of other members, and his behaviour will reveal a norm which is incompatible with that held by other members. If the network is nodal, the impact depends critically upon whether the new member replaces the existing nodal member or is a passive participant. Evidently, if the new participant is nodal, the potential for disruption is considerable.

Conversely, the new member becomes exposed to a belief set which is incompatible with that which he holds. As set up, his is exposed to a coordinated challenge to all three components of his beliefs. This increases the chance that the local equilibrium maintained by the interactions of the beliefs within his thought system will be dislodged.

Hence, as a result of potentially disruptive new membership, the psychological forces of *conformity*, *imitation* and *expression* may be sufficiently strong to overcome the forces reinforcing beliefs within individual participants, thereby restoring common beliefs across the network. Potentially, either the existing members or the new members may be the ones to adjust their beliefs depending upon the number of new members and whether any of them are nodal. Note that there is no scope for hybrid beliefs to be established, since all hybrids would be internally unstable.

However, an alternative process by which the network might restore compatibility of beliefs among its members is if it restricts its activities to those which do not generate incompatible interactions. For example, suppose that instead of the beliefs of the new member being entirely incompatible, only two components, say identities and norms, are so. If the network now restricts itself strictly to the circulation of narratives it can restore compatibility among its membership. Networks that do this evolve from being general purpose social clubs to a specialized core function.

How beliefs influence network selection

Above I left network participation as being exogenously determined. Initially members happened to have common beliefs and new members with incompatible beliefs then joined. In some contexts this is an entirely reasonable characterization. However, it is also evident that in many contexts network participation reflects the compatibility of beliefs. This is, indeed, the second form of *filtering*. A direct corollary of the taboo theory of Benabou and Tirole is that by participating in a network that is compatible with existing beliefs rather than one which is incompatible, many potentially disruptive interactions will be avoided. The investment in identity will thereby be protected.

3.3 Cultures and Cultural Change

I have defined a culture as a combination of beliefs and networks. Necessarily, cultures are slow-changing and the above analysis suggests some rudimentary reasons why this might be the case. At the level of the thought processes of an individual, the interaction of different components of beliefs generates beliefs that are mutually supporting. At the level of the networks in which individuals participate, network membership and the beliefs of individuals tend to interact so as to generate common beliefs that are reinforcing. As a result, cultures are locally stable. That is, they persist unless disturbed by a shock sufficiently large to move individuals and the networks in which they participate to a new local equilibrium.

Evidently, there is some process by which an initial assignment of beliefs among individuals, and of individuals to networks, becomes a culture. The initial assignment will be in disequilibrium. Five parameters then determine the processes that lead to equilibrium. The first is the attractive and antagonistic forces among the beliefs held by individual actors. The second is the mapping from the composition and circulation of beliefs in a network onto the beliefs of its members: for example, reinforcing those beliefs held by majorities and challenging those held by minorities. The third is the mapping from individual beliefs to network selection: for example, actors may sometimes face a trade-off between the value of the network and its compatibility with their own beliefs. The fourth is the generation of narratives: both those generated by particular members of networks and which circulate within them, and those which enter networks having been exogenously generated and which are potentially disruptive. The fifth is the mapping by which potentially disruptive information is filtered by networks.

Once beliefs are in equilibrium they become inputs into behaviour. Recall that identities map into preferences; narratives map into the causal relationships assumed to determine the consequences of behaviour; and norms impose internalized constraints. Given these mappings, the actor then chooses those behaviours which maximize utility as in standard economic analysis.

Feedbacks from Consequences of Behaviour

In a conventional rational choice framework the consequences of behaviour feed directly back onto decisions. Identities, norms and networks do not enter the analysis, and narratives are either unbiased estimates of true theories, or revised through a Bayesian process that gradually leads to true theories. In contrast, in the above setup, people are to an extent locked into the beliefs which they hold, and networks in which they participate, even if the consequences of the behaviour which their culture generates are adverse. People do not acquire a capacity for rational decision taking until long after they have adopted a set of beliefs (Hood, 2014, p.70).

The theory of taboos provides one explanation as to why feedback from adverse outcomes onto identities, narratives and norms may be too weak to force a change of identity. Setting the theory of taboos in the beliefs-networks framework sketched above may make it easier to test empirically. I have suggested that adverse outcomes should generate potentially disruptive news, which will be transmitted predominantly by means of narratives. As such, the passage from the raw news of the consequences of an action, to narratives that are received by the originating actor, will be filtered. Filtering will occur both by the suppression and distortion of news by the narrative generators within some networks and by the choice of network participation by the actor. Indeed, the central message of Haidt (2012), supported by copious and disturbing evidence, is that norms are relatively impervious to potentially disruptive information because inconvenient facts are effectively neutered by selective use of other evidence.

In effect, beliefs and network participation are in locally stable equilibrium. They are therefore liable to be undisturbed by the small flows of information generated by adverse consequences of behaviour. That is, cultures can be traps. How then do cultures change?

The interaction of material incentives and normative incentives

One potential process is through the introduction of material incentives which align with desired new norms and shift equilibrium behaviour: potentially, the material incentives need only be temporary to have persistent effects. Besley, Jensen and Persson (2014) analyze this interaction of material incentives and moral norms to see whether they are complements or substitutes. Their theory implies that there is considerable scope for material incentives to complement norms. Applied to tax evasion, if enforcement is intensified, so that the material penalty from the risk of prosecution for non-payment is increased, fewer people will evade tax. Because of this, the gap between the average moral intensity of those who pay and those who do not pay widens. As a result, choosing to be a non-payer reveals that the person is worse relative to the peer group than when penalties were lower. This in turn increases the shame from the decision to be a non-payer. More people choose to pay: the material incentive strengthens the power of compliance with the norm and so is a strategic complement. This is indeed what they find in the data. Due to random political events, councils periodically enforce tax payment more vigorously by increasing the prosecution of non-payers. This generates not only a direct effect on compliance, but also a favourable interaction effect.

While the above analysis of the interaction of beliefs does not call into question this interpretation, it does suggest a further reason to expect a positive interaction. The action of increasing the prosecution of tax evaders generates a narrative that 'tax evaders will be punished'. As indicated by the link between norm and punishment suggested by Robert Akerlof, this reinforces the norm that 'evading tax is wrong'. The tougher narrative is compatible with a tougher norm. In terms of the BJP analysis, tougher material penalties directly increase the mean intensity of the moral commitment to taxpaying, so that the guilt from non-compliance increases.

However, material incentives can also be strategic substitutes for the moral authority of norms. The famous case of the introduction of fines for late attendance at an Israeli nursery illustrates such a case. Inadvertently, the introduction of a fine generated a new norm that 'you should bring your child on time unless you pay extra'. This diluted the previous moral norm 'you should bring your child to school on time'. Parents could potentially have interpreted the new fine differently: as 'bringing your child to school late generates such inconvenience for others that to remind you of it we will impose a token fine'. But as implemented, parents were left free to choose whichever interpretation which was most convenient to them. The 'Haidt bias' ensured that the network would filter the event by generating the former interpretation. Thus, the error made by the nursery was not to pre-empt such filtering by using its power as the nodal actor in the network. It should have presented the fine as being a *token reminder* of an existing norm: the

considerable inconvenience to others. Instead, the fine was self-servingly interpreted by parents as *quantifying* that inconvenience and thereby providing a sufficient compensation for it.

The two distinct interaction effects by which material incentives change the shame of being relatively bad, and the guilt of being absolutely bad, will generally both be present. In the Israeli nursery case not only did the fine weaken the sense of guilt, but by reducing compliance, it reduced the moral gap between compliers and non-compliers and so weakened the sense of shame. Note, however, that in this case the shame effect can only arise as an interaction with the guilt effect which is necessarily primary.

Catastrophic change

One process by which equilibrium is dislodged is if the outcomes consequential upon the culture suddenly deteriorate for a majority of the participants in the network. The sudden deterioration may be genuinely consequential on a persistently dysfunctional culture: for example, debts may accumulate to a point at which the country ceases to be creditworthy, generating a sudden large fall in the standard of living. However, the sudden deterioration may be exogenous to the cultural dysfunction but apparently attributable to it. In either case, it is the common experience of a sudden large deterioration that matters. That is, it is not sufficient that outcomes for everyone are highly adverse, nor that outcomes deteriorate massively for a minority. Either of these situations can persist as the equilibrium because they do not generate a coordinated shock of new disruptive information that overwhelms cultural defences.

An example of a coordinated adverse shock that produced cultural change is the defeat of Germany in the Second World War. This has generated a profound shift in beliefs about warfare: modern Germany is famously pacific, in contrast to the century prior to 1945. The common shock of defeat provided the necessary coordinated challenge to existing beliefs. The shock also radically disrupted the social networks that had supported prior belief systems. In contrast, severe social dysfunction in Africa has proved to be highly persistent. This is consistent with the theory of cultural persistence because either there has been no moment of common catastrophe, or it has been neutralized by a narrative that attributes liability to some malevolent external actor such as the IMF.

The other process of disruptive change is if there is a massive, coordinated change in beliefs flowing into the networks of the culture independent of the outcomes which the culture generates. This can occur either because external networks become newly attractive to members of the culture so that many people choose to join these networks *en masse*, or because the external networks in which members already participate as a minority undergo so radical change of beliefs.

An example of such an externally induced process is the remarkably rapid change of beliefs in the American South during the 1960s following a century in which beliefs were stable. A plausible explanation for this change is the cultural changes that occurred within American youth

culture during the 1960s. Not only did beliefs change among American youth, but norms were added to networks which had been norm-light. This provided a massive coordinated shock to the existing Southern culture which it could not withstand. Quite possibly, the new technology of global networking that developed in the first decade of the 21st Century may help to explain the Arab Spring. Not only was the network useful domestically for the coordination of protest, but it may have shifted beliefs among Arab youth so that they became radically incompatible with the belief system of the prevailing culture. The traumas of the ensuing turmoil may have constituted a common large deterioration such that the suspended beliefs that had been held by the older generation were restored.

Each of these processes depends upon massive injections of disruptive information that overwhelm filtering mechanisms. However, there are also more incremental ways in which cultural change can occur and these may be more pertinent for practical policy.

If the starting point is that beliefs are strongly held, with values near to saturation, then a direct assault through exposure to disruptive beliefs is likely to fail. An alternative approach is first to weaken the intensity of beliefs by reducing the flow of reinforcement generated by networks. Either the rate of activity of the dysfunctional networks can be reduced, or the participation of the target group of actors in them can be reduced. This process takes time, as the rate of decay gradually establishes a new equilibrium in which beliefs are less strongly held. Once this equilibrium has been reached, the coordinated disruptive information may then succeed in overwhelming it.

3.4 The behaviour of principals

In the above setup, actors do not behave strategically. For example, they do not join networks with a view to changing the equilibrium beliefs of the network, nor do they generate or distribute narratives with a view to their ultimate consequences for behaviour.

In contrast, principals aim to alter the equilibrium behaviour of actors so as to achieve some preferred outcome. For example, at the level of society, the principal might be the government, attempting to maximize either the personal interests of the political elite, as in public choice theory, or the interest of citizens, as in social choice theory. The principal uses control variables to further his objectives. Within the above framework, these control variables can only be drawn from among the initial endowments and the five dynamic forces set out above.

A reasonable restriction on the range of control variables available to the principal is to exclude mind control: the beliefs held by actors cannot be directly controlled by the principal. The influences of government on citizens, of CEOs on workers, and of parents on children all generally fall far short of this. The exclusion of mind control also precludes three of the five forces of adjustment from being control variables. The adjustment of incompatible dyads of beliefs; the adjustment of individual beliefs in response to participation in networks with

different beliefs; and the network participation decision; all occur directly within the minds of individual actors.

This leaves three potential control variables. First, the principal will usually be able to alter the endowment of networks, create some and suppressing others. Clearly, to be interesting, these powers cannot be absolute. Second, the principal will be able to generate external narratives which can then be injected into networks, although they will be subject to the filters that the network operates. Finally, the principal may be a nodal actor in some networks, and thereby able to filter or neutralize those narratives which are potentially disruptive in respect of the objectives of the principal and perhaps be able to influence identity as a role model.

In the structure described above, economic behaviour remains the result of decisions which maximize utility, as in conventional economic analysis. But the entire decision process is subject to preferences, perceived causal relationships, and internalized constraints on choices which are the outcome of interactions between beliefs and networks that have generated a locally stable cultural equilibrium. Economics is not wrong to insist on individually maximizing behaviour: this is both a powerful insight, and a valuable discipline on analysis. But individual maximizing behaviour is only the final stage of a more complex, but still manageable, social psychological process. This is the necessary base from which to investigate the scope for principals to use the three control variables. That is, the options for the principal can be analyzed as attempts to shift actors between equilibria.

4. Three Levels of Behaviour

Cultures apply to a particular sphere of behaviour. There are, for example, dysfunctional families in functional societies: citizen behaviour is locally stable in a good equilibrium while family behaviour is stuck in a locally stable bad equilibrium. The beliefs applicable to each sphere of behaviour interact with networks that are also specific to each sphere. I now turn to three distinct spheres of behaviour: people as citizens, workers and family members.⁸

Having a common framework for apparently very different behaviour has the potential for cross-fertilization. However, the most important consequence is that in each case it introduces another actor who can be viewed as the principal in a principal-agent game with people whose behaviour is shaped by the processes discussed above. At the level of the person as citizen, the principal is the government. For present purposes I simplify by taking the objective of government to be the optimization of public goods provision. At the level of the person as worker the principal is the manager of the firm, and at the level of the family the principal is the middle generation between the young and the old. The introduction of these principals grounds the analysis in a clear framework of maximization of an objective under constraints by means of control variables. In each case the control variables available to the principal are the identities, narratives and norms

⁸ The same structure could be applied to people as consumers, with firms as the principals influencing identities, narratives and norms through advertising.

that shape people's behaviour, and the networks through which they acquire them. The principal does not fully control these variables but has some influence over them.

4.1 Behaviour as a citizen

The outcomes of behaviour as citizen are the provision of public goods and transfers for social protection; outcomes differing markedly between societies. Some differences are due to historical divergence and others to the efforts of government as principal to shape beliefs.

Differences in mutual regard

An example of historical divergence is 'mutual regard'. Such regard is generated by the interaction of identity with a cluster of norms. By 'regard', I mean a presumption to trust others; a willingness to cooperate with others; and a feeling of sympathetic generosity towards others. Such norms in themselves are not necessarily useful for those who hold them: indeed, in some contexts they could be suicidal. By 'mutual' I mean, straightforwardly, that these attitudes are reciprocated by their beneficiaries and so become socially functional. The process by which norms become reciprocal is conditional upon the emergence of a shared identity. A sense of common identity does not in itself generate mutual regard: the norms intrinsic to the identity might just as well be self-sufficiency (Montana), or mutual suspicion (Sicily), as mutual regard. Hence, mutual regard is neither simply a specific type of norm, nor a specific structure of identities. Nor, despite its social value, is it a primary analytic building block. It is generated by the interaction of a specific cluster of norms and a specific structure of identities.

Mutual regard makes the successful more willing to finance transfers to the less successful. Although such transfers have become heavily politicized as a conflict between ideologies, they are more truly rooted in how people regard each other. By this I do not mean how the welfare of other people anywhere on Earth should be counted, as in the Universalist version of Utilitarianism, but how we regard other members of our own society, and by extension, how we define the limits of what we recognize as the society to which we belong. Mutual regard, or sympathy, gives rise to feelings of loyalty and solidarity for those fellow members of our community who are less fortunate.

Mutual regard enhances cooperation. Cooperation is enhanced by trust, but to be other than quixotic, trust must be underpinned by a reasonable presumption that it will be reciprocated. The bedrock of rational trust is knowledge that the society is characterized by mutual regard: because people have some sympathy for each other, it is sensible to presume that a cooperative action will be reciprocated.

Through research in experimental economics we now understand what enables cooperative outcomes to persist. In the vicinity of the fully cooperative outcome, each individual has a strong incentive to free-ride, so cooperation should usually be unstable. The persistence of cooperation turns out to depend on more than just widespread benevolence. The vital ingredient is that there

should also be sufficient people who punish those who do not cooperate. Punishment is costly, so people will only be prepared to do it if they have sufficiently internalized not just benevolence, but moral outrage at those who free-ride. Cooperative outcomes are fragile because if enough people back away from punishment then non-cooperation becomes the rational strategy. The role of hero performed by the people who punish non-cooperation in turn creates the possibility of ultimate villains. The minor villains are the people who do not cooperate, but the super-villains are people who punish the heroes. Again, since punishment is costly, systematically to get satisfaction from punishing the heroes can only arise if some people feel moral outrage not against the people who undermine cooperation, but against the people who try to enforce it. Why might some people have such dysfunctional moral codes? Some people might feel an overriding personal loyalty to others even if they free-ride, and so are outraged by those who punish them for doing so: identities have become too powerful relative to norms.

If the level of trust differs markedly between societies, the tactics people adopt in games that require cooperation will also differ. This has recently been tested through experimental games (Gaechter *et al.*(2010)), the same game being played under standard laboratory conditions by university students in sixteen countries. In some societies super-villains appear to be ruinously numerous. If heroes punished another player for free-riding, the outraged response was to punish the heroes. The researchers then investigated whether these differences in behaviour were related systematically to observable characteristics of the countries in which the students lived. Directly, the differences in behaviour were related to differences in social capital, in other words to trust. But these in turn could be related to differences in the rule of law. In countries where the rule of law was weak people were opportunistic and so untrusting, and were inclined to be super-villains in cooperation games. I suspect that these differences in the rule of law can be traced yet further back to the difference between moralities based on loyalty to the honour of the clan, and moralities based on the Enlightenment concept of good citizenship. Super-villains should have a bad conscience according to the standards of the Enlightenment, but they are behaving morally according to the precepts of loyalty to the clan.

Government influence on beliefs and networks

The entity with the power of agency to affect these outcomes is government. Government is the principal in the citizen game, aiming to optimize public goods provision. Its control variables are the identities, narratives and norms that citizens adopt. Although each of these has an exogenous component, government can influence them through various policies.

Influence on identities

It is well-established that the social composition of identities matters for cooperation and generosity: diversity makes cooperation and the acceptance of transfers more difficult. Mismatches between the structure of identity and the structure of power are therefore problematic.

For example, in Africa identities tend to be predominantly sub-national and supra-national, whereas power tends to be highly concentrated at the national level.

Over time, government policies can affect the degree of diversity in society in various ways. Diversity can be reduced by building a sense of shared identity. As discussed above, Miguel (2003) has shown that the contrasting approaches to identity of President Nyerere in Tanzania and President Kenyatta in Kenya led to long term consequences for the ability of these societies to cooperate in public goods provision. Nyerere recognized the scope for, and economic importance of, using the power of leadership to influence the identities that people adopted.

Influence on narratives

The Charlie Hebdo assassinations generated potentially disruptive information across Moslem societies. The manifestly immoral act challenged the support that widely shared moral norms provided for the oppositional identity of radical Islam. Those political leaders who depended upon oppositional Islamic identity faced a dilemma: condoning the killings would violate widely shared norms, but condemning them would weaken the oppositional identities which they were promoting. The public response of the Mayor of Ankara is an example of a nodal political actor resorting to a neutralizing causal proposition: his comment was that he believed the CIA to be implicated in the killings. This new causal account, which did not have to be fully believed in order to be effective, protected the mutual support of identity and norms.

Influence on norms

Natural resource exploitation raises normative issues of public policy: to whom should the benefits accrue, both spatially and temporary? In Botswana norms were pre-emptively shaped by the president, Sir Seretse Khama. At the time of prospecting, he emphasised that clan-based ownership rights would cause vast inequalities between clans, whereas collective ownership would preserve inter-clan equity. Clan leaders thereupon chose collective ownership of any discoveries. To protect the interest of the future he introduced the mantra 'we're poor and therefore we have to carry a heavy load', providing a simple rationale for patience. These seemingly trivial communication strategies had dramatic long term consequences, which can be seen by contrasting the experience of Botswana with that of Britain, Nigeria and Tanzania. In these three countries no pre-emptive attempt was made to make the ethical case for spatially shared ownership. In all three the Haidt bias, by which people come to see as righteous norms which are in their self-interest, generated powerful local claims to ownership. In Nigeria it led to a civil war, in Tanzania to local riots, and in Scotland to political secession built around the slogan 'its Scotland's oil'. In contrast, in Botswana the fairness of national ownership of diamonds has never been challenged. Similarly, whereas the moral acceptance of patience enabled Botswana to build up a large Sovereign Wealth Fund, in Nigeria, which was more typical, the big increase in oil revenues in 1974 was followed the next year by a 75% public sector wage increase.

Public norms also affect the propensities for cooperation and generosity. In South Sudan widespread public norms such as ‘the Dinka won the war and have a right to rule the country’ are less conducive to cooperation than those such as ‘liberty, equality and fraternity’.

Influence on networks

Governments can use networks to influence identities, narratives and norms. An instance in which the effects of network use by government have been precisely measured is that of the Rwandan radio station *Milles Collines* by the Government of Rwanda in 1994 to promote Hutu killing of Tutsi (Yanagizawa-Drott, 2012). In this case both a narrative and an identity were transmitted.

Another salient example is Russia following the Orange Revolution in Ukraine. Reputedly, President Putin recognized that if such a revolution could occur in Kiev it was also possible in Moscow. The diagnosis of the KGB was that the revolution was the result of social networks organized by NGOs. Rather than suppressing NGOs, the KGB counter-strategy was to build a rival set of government-organized NGOs (‘GONGOs’) and to build the capacity for sophisticated news management in the media. Currently, this is clearly successful in shaping Russian perceptions of events in Ukraine.⁹

However, the most striking examples of the use of networks in government are the mass national parties: the Nazi Party in Germany, the Communist Parties in China and Cuba, the ruling parties in Turkey, Ethiopia and Rwanda, and the briefly empowered Moslem Brotherhood in Egypt. While the ideologies of these parties differed considerably and some were disastrous, they had in common the technique of mass membership combined with a continuous program through which recruits were induced to internalize a set of core narratives and norms. The nationalist cast of these parties enabled them to provide a privileged ‘vanguard’ identity to party members, while at the same time offering an inclusive identity to the majority of the population. In important respects these parties have been remarkably successful. For example, Cuba has succeeded in creating a large cadre of highly motivated doctors without the need to pay high salaries. This would be difficult to understand in terms other than successful internalization of norms. A mass national party in which many members have internalized a common set of values and beliefs gives the government a powerful means of achieving its objectives not available to a democratic government.

4.2 Behaviour as a worker

I now turn from people as citizens to people as workers. The outcome that is determined by behaviour as a worker in an organization is productivity. High productivity is socially challenging because technologically it usually requires scale and specialization. However, each

⁹ I am indebted to Prof. Garton-Ash for this information.

of these introduces adverse individual maximizing behaviours. Scale makes it harder to maintain worker motivation; specialization introduces the need for coordination.

The principal is now no longer the government but the top management of the firm. It has considerable powers of agency and differences in productivity are likely to be attributable predominantly to current and past management decisions. Economists have focused almost exclusively on incentives, an approach which has gradually permeated much of the public sector. However, in the private sector firms appear to take a broader approach, trying to influence the identities, narratives and norms that workers adopt.

Influence on identity

The potency of influencing identities has been a key theme in the work of Akerlof and Kranton (2011). They show that effective organizations persuade their workers to internalize the objectives of the firm: to become ‘insiders’. As an illustration of this effectiveness, here is the story of one of my early students, a radical Marxist, who somewhat reluctantly accepted a job in Unilever where he found himself a member of a team marketing soap powders. Two years later he visited me. After recounting with some passion the challenges posed by this marketing, he remarked ‘Of course, I couldn’t possibly work for Procter and Gamble’. In effect, in only two years Unilever had succeeded in turning this somewhat unpromising recruit into an insider: the proverbial company man.

Influence on narratives

CEOs, like political leaders, can also seed narratives: stories which explain an aspect of the environment facing workers. For example, here is a narrative promoted by the CEO of the highly successful company South-West Airlines. ‘The new United shuttle from LA to Dallas is an Exocet aimed at the heart of South-West Airlines’.¹⁰ The purpose of this narrative was evidently to motivate through fear of competition, and perhaps also to build oppositional identity.

Gibbons and Henderson (2013) demonstrate that even within narrowly defined categories of activity there are large and persistent differences between firms in worker productivity. They attribute this to differences in the ability of management to inculcate ‘relational contracts’ with their workforce. A relational contract is one which is insufficiently precise to be enforced through a court, but still provides guidance for workers as to how they can expect management to behave towards them, and how in return management expects them to behave. It is a contract, but not one that can be legally enforced. The vehicle for building and sustaining relational contracts is narratives: ‘*the prevalence of story-telling in organizations can be seen as an attempt to promulgate prototypical examples*’ [of behaviour]. (Gibbons and Henderson, p.60).

Influence on norms

¹⁰ Harvard Business School case study.

CEOs can also inculcate norms, encapsulated for example in their mission statements. John Kay has written persuasively that the mission switch of ICI senior management from an aspiration to be the best chemical company in the world to one of ‘maximizing shareholder value’ marked the decline in the company’s fortunes (Kay, 2010).

Mayer (2013) analyses how firms can inculcate norms of commitment to various types of behaviour. Because contracts can never be complete and are costly, it is efficient to reduce the cost of transactions through making commitments to various norms. CEOs make a multitude of commitments that their firm will not exploit time-inconsistency. They attempt to reassure workers that investing in firm-specific skills is worthwhile; to reassure customers that the firm will maintain after-sales responsibility for products; and to reassure society on sustainability (for example, the BP campaign ‘beyond petroleum’). However, to be effective these commitments must be credible. Mayer discusses how firms can make commitments credible both by internal procedures for accountability and by structuring ultimate authority so as to give these interests some power over decisions. For example, by having workers on oversight boards, German firms have been able to make credible commitments to workers that firm-specific training will not be exploited.

An application: The African public sector

While in high-income countries private sector firms have necessarily had to learn how to motivate their workforce through approaches beyond incentives, in contrast, much of the public sector in poor countries suffers from very low productivity despite reaping the benefits of scale and specialization. In effect, top management has not succeeded in building effective organizations.

For example, In Uganda schoolteachers actually teach for an average of only two hours per day, despite being paid for seven. Only four percent of Ugandan maths teachers are able to gain 80 percent on the standard tests set to their students. In Nigeria, a substantial proportion of teachers of ten-year olds are functionally illiterate. In the health sector the theft of drugs by nurses is usually endemic, but poor conduct extends right up the hierarchy. For example, in Zambia, the Permanent Secretary of the Ministry of Health set up a company to exploit the offer of the Global Fund to pay for anti-retroviral drugs, using his official position to authorize the purchase of drugs imported by his company. Worse, he chose to import fake drugs to increase his margins. Nor is this poor performance simply a reflection of inadequate pay. Barr et al (2009) in a tracking study of Ethiopian student nurses, show how norms of good conduct prevalent upon completion of training are eroded once they are placed as junior staff in public clinics where norms among senior staff are corrupt.

A study by Reinikka and Svensson (2010) compared public health clinics with mission-run health clinics and found that the latter had considerably higher performance despite significantly lower pay levels. This coexistence of ineffective and effective organizations in the same society

is widespread in Africa. For example, in Anambra State of Nigeria the Governor handed over the management of many secondary schools to local churches (while maintaining public funding), in recognition that in the past the churches had been able to oversee effective schools.

A final example is the difficulty of building a tax system in DRC. As usual, the IMF recommended the introduction of VAT, which was duly adopted. However, in DRC (as in several other African countries) the VAT is a revenue-*loser* because the rebate component of the tax is being exploited by fake receipts, whereas the payment component is being evaded. Evidently, the tax administration has not been able to create a belief system in which tax inspectors internalize the objectives of the organization. While they may have internalized an identity 'I am a good tax inspector', they have evidently not internalized the norm 'a good tax inspector is one who raises revenue for the nation'. More likely, they have internalized a norm such as 'a good tax inspector uses the opportunity to extract bribes to help his family'. The norm and the identity are supported by narratives such as 'If you pass revenues on up the tax system somebody else will steal them', and 'If you help your family, they will help you'. This is a locally stable belief system because the identity, narratives and norms are mutually supportive.

Policy for reforming African public services has focused overwhelmingly upon incentives. Yet the above problems appear to reflect deeper failures of internalization of identities, narratives and norms. For example, the teachers whose behaviour is described above cannot have taken what Akerlof and Kranton regard as the key step in worker identity: 'I am a good X'.

I have argued that in such situations governments should fund effective non-government providers such as mission schools, enabling them to expand their activities (Collier, 2007). This is based on the premise that it is easier to scale up those organizations in which the workforce is already motivated, than to motivate workers in unmotivated organizations. The rationale for this follows from the network property of sequence-dependence. When a motivated organization is scaled up, recruits are added sequentially to an existing stock and so are likely to adopt the norms of the existing network. An important issue for research on this topic is the extent to which the healthy norms of workers in functional organizations are achieved through the careful selection of recruits predisposed to the norms, relative to imitation and conformity post-recruitment.

Since most public services will continue to be delivered directly by the state, it is also essential to learn more precisely quite why, and in what respect, the beliefs of the workforce are so dysfunctional and how this could be rectified. In the 1990s tax collection was reformed through the creation of Independent Revenue Authorities. This involved closing the existing tax collection agency and re-recruiting a minority of screened workers into a new organization with a higher level of pay and new narratives designed to create new norms. While there has been some work by the IMF evaluating performance, such experiments have the potential for detailed analytic anthropological study of how the government's attempt to use new narratives and networks as control variables affected the identities and norms of the workforce.

Prior to reorganization, such a study would start by establishing the distribution of salient identities, narratives, norms and networks over the workforce. It would assess the likely stability of these beliefs by measuring their intensity, the extent of their interdependence, reinforcement through network participation, and network filtering of disruptive information. The cultures of workers would then be related to their measured performance. During the reorganization, the study would measure the use of the non-material control variables - changes in networks and narratives. Post-reorganization it would re-measure beliefs, networks and performance and track them over sufficient time to establish their stability. Ideally, there would be sufficient variation both in cross-section between workers, and over time as a result of the reorganization, to establish clear causal relationships between beliefs and performance, and between networks, disruptive narratives and equilibrium beliefs. From this base it would be possible to analyze the interaction between culture and material incentives which is the ultimate objective of an economic analysis of culture.

4.3. Behaviour as a family member

Finally I turn to people as family members. The outcome on which I focus is non-market inter-generational transfers although other aspects of family-oriented behaviour might also be important. Caring for the young and the old is not well-suited to purely market provision. However, it is socially problematic because the basic inter-generational deal – the receipt of care when young in return for the provision of care when an adult – is not time-consistent.

The power of agency here lies with parents. Parents can attempt to inculcate an identity based on family, a narrative of success attributable to intergenerational assistance, and norms which depict inter-generational responsibility as honourable. Clark (2014) analyzes rare surnames and generates astonishing evidence of the persistence of family success across generations. He concludes that this is either genetic or cultural.

Influence on identity

The intensity of family identity varies widely, with an apex in dynastic ambitions common in successful political and business households contrasting with a nadir in which children do not know their parentage. Clark gives evidence of the use of symbols such as heraldry and estates to create dynastic identities. The *WDR2015* has a good analysis of the consequences of the other end of the spectrum of family cohesion, exemplified by Romanian orphans.

Influence on narratives

Probably the most common of all narratives are family sagas: such stories, replete with black sheep and self-sacrifice, both reinforce identity and illustrate the consequences of choices of care. Stories can also illustrate norms, though the norms themselves are generic. Parents can either reinforce or repudiate these norms through their provision of care to younger and older generations.

Influence on norms

Perhaps the most fundamental revolution in norms is the passage from codes of honour to codes of justice (Pinker, 2011). Maintaining honour is an obligation of the family, maintaining justice is an obligation of the state. Although family identity and national identity are entirely compatible, the obligation to family honour is not compatible with the obligation to justice: when a family member breaks the law, these obligations collide. Hence, the emergence of the rule of law depends upon the erosion of the norm of honour. In turn, as discussed above, the rule of law appears to be critical for the emergence of trust beyond the family.

Influence through networks

Clark's evidence on the transmission of family success across generations is at least suggestive of the transmission of culture through a family network. However, recent evidence from child psychology suggests that children are more strongly influenced by the behaviour of peer groups than by that of their parents. Indeed, parental use of rewards and punishments to induce desired behaviour can even prove counterproductive (Wilson, 2011). However, to an extent parents control the peer groups with which their children interact. Hence, as argued by Judith Rich Harris, parents can still influence the behaviour of their children at one remove, via the selection of peer groups (Harris, 1998). If my interpretation of the evidence on child behaviour is correct, this is a clear instance of where the approach of influencing behaviour through the identities, narratives and norms that are accessed through networks is more effective than the conventional economic approach of incentive design.

Wantchekon (2014) provides striking evidence from Benin of the power of family and peer group networks in inducing imitation. The choice on which he focuses is whether to send children to school. The initial provision of education in the interior of Benin was through an essentially random assignment of missionary teachers. But this initial random assignment then produced powerful inter-generational differentiation as children followed role models from the previous generation. He is able to demonstrate the diffusion of role models beyond the family to other families which were part of the same network as a result of spatial proximity.

5. Conclusion: the Implied Research Agenda

The proposed structure opens several possibilities.

The analysis provides a clear and testable account of why dysfunctional cultures are locally stable despite generating observably bad outcomes and of what to do about it. There are eight distinct points of intervention. Three concern networks: the filtering activities of nodal principals might be disrupted; neutralizing propositions might be discredited; participation in the social network might be disrupted by encouraging participation in another network in which beliefs circulate that are incompatible with existing beliefs. Five concern the mutually reinforcing nature of belief systems. Unbundling these interdependent beliefs into identities, narratives and norms

may explain why soap operas have been found to be so effective in changing behaviour, (for example, viewing soaps in which there are many divorces increases the divorce rate). A soap opera works through five routes. First, the direct observation of the norm-based behaviour of TV characters provides new information to the viewer as to what other people will be seeing as 'normal'. This changes the signal value of analogous behaviour: e.g. 'divorce is pretty common'. Second, the direct observation of the apparent consequences of actions, and any narratives about the consequences of actions related by characters, provide new information about causal relationships. E.g. 'divorce makes you happier'. Third, identification with sympathetic characters as role models directly changes preferences: e.g. 'I'm not the sort of person who stays married to a loser'. Fourth, the soap opera generates common knowledge, which has been found to be critical for achieving coordinated change (Thomas *et al*, 2014). Fifth, precisely because it works through each of these four routes, a soap opera constitutes a 'big push' approach. Since identities, narratives and norms are mutually reinforcing, the fact that all are being pressured together enhances the changes of breaking out of the locally stable equilibrium. The effects of these distinct routes can potentially be quantified through standard methods.

The analysis clarifies how framing works. When primed to increase the salience of their professional *identity*, bankers are substantially more likely to cheat at a coin-tossing game (Cohn *et al*, 2014). Directly, priming works because of the effort required to process information. Not all stored information can be held at the front of the mind at the same time, so that priming rearranges the ordering of awareness of identities. However, the study is only of interest because the increased salience of this identity 'dragged along' with it a norm *which was not directly mentioned in the priming*. Evidently, particular norms 'fit' with particular identities, while others are incompatible. Further, the fact that bankers as a group had a common reaction to the increased salience of their professional identity suggests that they shared this pairing of identity and norm, and this must have been derived from participation in a common social network. The different components of a system of beliefs are not only acquired but regularly refreshed from participation in a network.

Analogously, through a randomized controlled trial, Collier and Vicente (2014) studied the effect of increasing the salience of a *norm*, the context being a Nigerian election. During the campaign, the message 'Don't vote for violent politicians' was broadcast through street theatre, tea shirts and similar public awareness techniques. The message significantly altered voting patterns. While votes did not switch, the supporters of violent politicians became less likely to turn out, and the supporters of their non-violent opponents became more likely to turn out. A plausible explanation is that by raising the salience of a reasonable moral norm already held by those who heard it, this message accentuated the standard Augustinian tension between desires and morality. Abstention would be a natural solution to such a tension: a vote for the violent politician would be to breach the moral norm, while switching to the non-violent politician would be to act against preferences.

The analysis provides a psychological substructure for signalling, whereby revelation of ‘type’ changes the prediction made by observers concerning the behaviour of the actor. For this to happen the observed action must regularly be matched in a population with the unobserved characteristic which is revealed. Various attributes must therefore naturally cluster. The mutual attraction of equilibrium values of identity, narratives and norms provides an account of this clustering. Dysfunctional cultures trap individuals in damaging activities designed to signal their type (Seabright, 2012). By clearly measuring distinct clusters of beliefs we would be better able to predict which behaviour was incompatible with which belief set. These incompatibilities are what generate the signal.

The analysis also suggests a research agenda.

Beyond the conventional analysis of material incentives, behaviour can be separated into three fundamental relationships. The first is how choices are affected by beliefs. The second is how equilibrium beliefs are generated through interaction within networks. The third is how a principal can use some aspects of beliefs and networks as control variables to achieve an objective.

Evidently, the relationships form a hierarchy. Until the first stage is properly established, the second stage lacks consequence. Until the second stage is established, the third stage cannot be undertaken. However, subject to this hierarchy, each of these stages is sufficiently self-contained that it can be researched, both theoretically and empirically in isolation.

This three-stage structure applies for each of the three beliefs: identities, narratives and norms. It also applies in each of the three behavioural contexts: citizen, worker and family. Even with these nine channels of psychological influence the structure remains highly reductionist: the same three fundamental relationships should each be researchable in nine apparently very different contexts. For example, business schools investigating how HR practices affect worker internalization; sociologists investigating low literacy in dysfunctional families; and development economists investigating free-riding in public goods provision, are all studying the first relationship: how beliefs map into behaviour. Similarly, political scientists studying how Alex Salmond, leader of the Scottish Nationalist Party, has succeeded in changing perceptions of identity; business schools producing case studies of CEOs who successfully raised productivity without changing material incentives; and economic historians studying the inter-generational success of family businesses, are all studying the third relationship: how a principal can influence culture to achieve an objective.

It may be that within each relationship there are some common analytic properties and empirical regularities yet to be recognized. If so, once recognized they might make social science more integrated without resort to the intellectual imperialism by which economics has to date sought to occupy the behavioural territory of other social sciences without incorporating their insights.

References

Akerlof, G. and R. Shiller, 2009, *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*, Princeton.

Akerlof, G. and R. Kranton, 2011, *Identity Economics*, Princeton University Press.

Akerlof, R. 2013, Anger and Enforcement, mimeo, University of Warwick.

Alesina, A., Glaeser, E. and Sacerdote, B. 2001, Why doesn't The US have a European-style welfare state? Harvard Institute for Economic Research, WP 1933.

Barr A., M. Lindelow, P. Serneels, 2009, Corruption in public service delivery: an experimental analysis, [*Journal of Economic Behaviour and Organization*, 72 \(1\) 225-239](#)

Besley, T. J, A. Jensen and T. Persson, 2014, Norms, Enforcement and Tax Evasion, mimeo, LSE and IIES.

Benabou, R. and Tirole, J. 2011, Identities, Morals and Taboos: Beliefs as Assets, *Quarterly Journal of Economics*, 126, 805-855.

Candelo-Londono N. et al, 2011, Social Exclusion and Identity: a field experiment. Mimeo, University of Texas.

Cialdini, R., 2007, *Influence: the Psychology of Persuasion*, Harper.

Clark, G., 2014, *The Son also Rises*, Princeton University Press.

Cohn, A., E. Fehr, and M.A. Marechal, 2014, Business Culture and Dishonesty in the Banking Industry, *Nature*, 516, 86-69.

Colley, L. 1992, *Britain: Forging the Nation*, Yale University Press.

Collier, P. and A.J. Venables, 2014, Closing Coal, *Oxford Review of Economic Policy*.

Collier, P. and P. Vicente, 2014, Votes and Violence, *Economic Journal*.

Collier, P., 2007, *The Bottom Billion*, Oxford University Press.

Dijksterhuis, A. 2005, Why we are Social Animals, in *Perspectives on Imitation: from Neuroscience to Social Science*, Vol. 2, ed. S. Hurley and N. Carter, Cambridge.

Fisman R. and Miguel E. 2007, Corruption, Norms and Legal Enforcement, *Journal of Political Economy*, 115, 1020-48.

Gaechter, Simon, Benedikt Herrmann & Christian Thöni, 2010. Culture and Cooperation, CESifo Working Paper Series 3070, CESifo Group, Munich.

Gibbons, R. and R. Henderson, 2013, What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises, WP 13-020, Harvard Business School.

Greene, J. 2013, *Moral Tribes*, Harvard University Press.

Haidt, J., 2012, *The Righteous Mind*, New York, Pantheon.

Hamlin A. and C. Jennings, 2011, Expressive Political Behaviour: Foundations, Scope and Implications, *British Journal of Political Science*, 41, 645-70.

Harris, J. R. 1998, *The Nurture Assumption: why children turn out the way they do*, The Free Press.

Hood, B., 2014, *The Domesticated Brain*, Pelican.

Kahneman, D., 2011, *Thinking Fast and Slow*, Penguin

Kay, J. 2010, *Obliquity*, Profile Books

Kuran, T. 1989, A Theory of Wildfires and Protests, *Journal of Economic Perspectives*.

Mayer, C. 2013, Conceiving Corporate Commitment: Creation and Confirmation, forthcoming in J. Hill and R. Thomas (eds.) *Shareholder Power*.

Miguel, E. 2004, Tribe or Nation? *World Politics*, 327-62.

Mulleinathan, S., J.Schwartzstein, and A. Shleifer, 2006, Coarse Thinking and Persuasion, *Quarterly Journal of Economics*.

Omerod, P. 2012, *Positive Linking*, Faber and Faber

Pagel, M. 2012, *Wired for Culture*, Norton

Pinker, S. 2011, *The Better Angels of our Nature*, New York, Viking

Postlethwaite, A. 2010, Social Norms and Preferences, in *The Handbook for Social Economics*, J. Benhabib, A. Bisin and M. Jackson (eds.). Elsevier.

Reinikka, R. and J. Svensson, 2010, Working for God, *Journal of the European Economic Association*, 8, 1159-78.

Seabright, P. 2010, *The Culture of Strangers: a Natural History of Economic Life*, Princeton.

_____, 2012, *The War of the Sexes*, Princeton.

Seligman, M.E.P. 2011, *Flourish*, Simon and Schuster

Sen, A. 2006, *Identity and Violence*, Allen Lane

- Smith, J.M.H. 2005, *Europe after Rome: a new cultural history, 500-1000*, Oxford University Press.
- Spolaore, E. and Wacziarg, R, 2013, How Deep are the Roots of Economic Development, *Journal of Economic Literature*, 51, 325-69
- Thomas, K., O.S. Haque, S. Pinker and P. DeScioli, 2014, The Psychology of Coordination and Common Knowledge, *Journal of Personality and Social Psychology*, 107, 657-676.
- Wantchekon, L. (with M.M Klasnja and N. Novta), 2013, Education and Human Capital Externalities: Evidence from Colonial Benin, Working Paper, Princeton.
- Wilson, T.D. 2011, *Redirect*, Little Brown
- World Bank, 2015, *Mind and Society: World Development Report, 2015*, World Bank.
- Yanagizawa-Drott, D. 2012, Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide, Working Paper, Kennedy School of Government, Harvard.
- Zak, P. 2012, *The Moral Molecule*, Bantam
- Zak, P. 2014, Why Your Brain Loves Good Storytelling, *Harvard Business Review* (October).
- Zingales, L., L. Guiso, and P. Sapeienza, 2006, Does Culture Affect Economic Outcomes, *Journal of Economic Perspectives*, 20, 23-48.