# PART IV

## Modern tribes

The communitarian conception of the Open Society

# 21. Joshua's question

In the 1973 film *Magnum Force*, Clint Eastwood, as detective Harry Callahan, is confronted by a group of three police officers who have committed a series of vigilante murders. The group try to persuade Callahan to join them, stating, "Either you're for us or you're against us." Callahan, displaying the moral backbone that exists alongside his maverick tendencies, declines the invitation, saying, "I'm afraid you've misjudged me." It's not clear whether Callahan's reason for refusing was accurate: by the end of the film he kills every member of the vigilante gang. If that doesn't count as being *against them*, I'm not sure what does.

This way of framing the central relation between an individual and a group, that one must either be *for* or *against* the group, is as old as humanity itself. It is given clear statement in the Book of Joshua, which in the King James' translation reads:

> And it came to pass, when Joshua was by Jericho, that he lifted up his eyes and looked, and, behold, there stood a man over against him with his sword drawn in his hand: and Joshua went unto him, and said unto him, Art thou for us, or for our adversaries? (Joshua, 5:13, KJV)

Not all translations phrase Joshua's question the same. The Common English Bible has him ask, "Are you on our side or that of our enemies?" The Good News translation asks, "Are you one of our soldiers, or an enemy?" The Lexham English Bible asks, "Are you with us, or with our adversaries?" And the New Living translation cuts right to the chase and simply asks, "Are you friend or foe?"

What these competing translations of Joshua's question highlight is the relationship between *belonging* to a group and the possession of certain *intentions* or *attitudes*. Some translations stress membership (are you "with us" or "on our side") and other translations stress intention ("art thou for us"). The Good News translation mixes the two, in that it treats being "one of our soldiers" and

**How to cite this book part**:

Alexander, J. McKenzie (2024) *The Open Society as an Enemy: A critique of how free societies turned against themselves*, London: LSE Press, pp. 223–300. https://doi.org/10.31389/lsepress.ose.e. License: CC BY-NC 4.0

being "an enemy" as mutually exclusive, ignoring the fact that there can be enemies within. Underlying all of this is the idea that our fellow group members are *friends* and, hence, *for us*.

This brings us to the communitarian conception of the Open Society or, what I will treat as equivalent, the rejection of tribalism. What exactly does that mean? In his seminal work *Two Sources of Morality and Religion* – where the concept of the Open Society first appeared – Henri Bergson characterised the closed society as one where people viewed every interaction along the same lines as Joshua's question: are you with us or against us? For Bergson, this closure coexisted with hostility to outsiders:

> The closed society is that whose members hold together, caring nothing for the rest of humanity, on the alert for attack or defence, bound, in fact, to a perpetual readiness for battle. (Bergson 1935, p. 229)

The trouble with Bergson's characterisation of the closed society is that it cuts too crudely. While any group that "[cares] nothing for the rest of humanity" and is ready to do battle with outsiders certainly counts as a closed society, it's not the only kind of closed society. We can relax this condition, requiring only that the group's primary or predominant concern is with its own members, with the "rest of humanity" given lower priority. If we read Bergson figuratively, such that being "on the alert for attack or defence" means not just a physical attack but also a readiness to respond to a *perceived critique* of certain *beliefs or values* – possibly ones seen as *constitutive* of the group identity – we obtain a broader characterisation of a closed society.

Yet even under this broader interpretation of Bergson, something seems to be missing. A closed society isn't just a society whose members are ready to fiercely defend the group from external critique, even when the critique targets beliefs or values central to the group's identity. *That* attitude is compatible with communities committed to free and open inquiry, the pursuit of truth, and evidence-based decision-making, so long as those communities don't roll over immediately in the face of external critique. A closed society must also feature a degree of commitment to the group that borders on the irrational, is unresponsive to critique, or is resistant to revision in light of contrary evidence.

It is this combination of irrationality and recalcitrance that Popper identifies as the defining feature of his sense of the closed society, which he also refers to as the "tribal society". He writes:

> There is no standardized 'tribal way of life'. It seems to me, however, that there are some characteristics that can be found in most, if not all, of these tribal societies. I mean their magical or irrational attitude towards the customs of social life, and the corresponding rigidity of these customs. (Popper 1945, p. 184)

Combining the accounts of Bergson and Popper, we arrive at the following: a closed society is one (i) whose primary concern lies with its own members, (ii) whose members are ready to defend themselves against perceived threats – by which we mean not just *physical* threats but threats to certain beliefs or values, often seen as core to the group identity – and where (iii) the beliefs or values triggering the defensive response are rigidly held with a degree of irrationality.[1] From this, we may then trivially define an Open Society as one that is not closed.

This proposed definition is quite a mouthful, but it is useful for several reasons. First, it doesn't treat open or closed societies as being purely binary: societies admit degrees of closure, and this definition allows societies to be more or less closed. Second, it distinguishes between closed societies and cases of mere group loyalty. Take, for instance, football fans. Even if conditions (ii) and (iii) are met, it's unlikely that (i) will be met to the degree required to transform a group of football fans into a closed society. Recall that condition (i) relaxes the Bergson requirement that a closed society "[cares] nothing for the rest of humanity". The new requirement, a society "whose primary concern lies with its own members", allows trade-offs to be made between its members and nonmembers. For the majority of football fans, the new condition fails to obtain: an ardent Manchester United supporter will still come to the aid of a Liverpool fan collapsed on the street. But when the first condition *does* hold – and as the trade-offs increase in severity – we arrive at something other than the loyalty of football fans. At the extreme, we arrive at the phenomenon of the "Ultras", a type of fanatic where team loyalty alone justifies violent, thuggish behaviour against supporters of other teams (Jones 2017).

A third reason why the proposed definition is useful is that, although it is formulated in terms of closed *societies*, it applies to any group that has a strong sense of collective identity. Bergson spoke of closed societies because given his focus on morality and religion, that was the appropriate level of organisation on which to focus. Popper spoke of closed societies due to his interest in the ideological conflict underlying World War II – a conflict between alliances of nation states. These historical facts notwithstanding, I think that focusing on closed *societies* alone is too narrow. In the last few decades, we have seen increased polarisation *within* societies along a number of dimensions. This polarisation is accompanied by an inward turn, a closing off, which shares many features with the defining features of a closed society but among groups. And so we need to appreciate that contemporary societies are increasingly composed of groups sharing a commitment to certain beliefs or values, which constitutes a group identity. These are the *modern tribes*. But the interactions between modern tribes are still implicitly governed by how they answer Joshua's Question.

In what follows, we will explore how the growth of modern tribes threatens the communitarian conception of the Open Society. To begin, there is the preliminary philosophical question as to what exactly we mean by "polarisation", and then there is the empirical question of whether polarisation has, in fact,

increased in contemporary society. After that, we need to ask how and why increased polarisation might matter. This requires us to engage with theories of social identity and how membership of a group can influence individual decision-making simply by virtue of belonging to a group as well as through the explicit pressure of social norms. Then we need to examine the underlying mechanisms that drive us to increasingly identify with tribes, which I think are deeply connected to concepts of authenticity and ideological purity. One concern about these drivers is how both can be exploited for political purposes, especially in the age of social media, returning to a concern we broached in Chapter 14. I then explore how the communitarian Open Society is threatened with fragmentation as we focus on what divides us instead of what unites us. Then I shall turn to the phenomenon of epistemic closure and how it relates to the formation and persistence of extreme groups. And, finally, we end with the question of what it means to live in an open society defined by diversity and difference of opinions and how to negotiate the collision of horizons this yields.

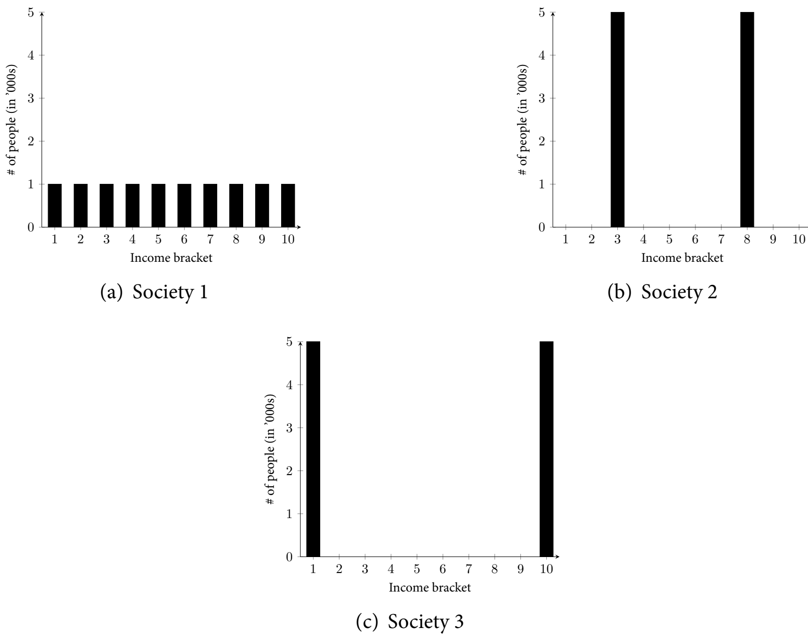
# 22. On polarisation

Polarisation exists when a population is divided into two or more groups with strong differences between the groups. This intuitive idea can be made more precise by drawing on the work of Esteban and Ray (1994), who studied measures of income polarisation. (We'll be interested in many types of polarisation, but polarisation of income is a good place to start because incomes, unlike ideas, are more easily measured.) As a starting point, they proposed such a measure should satisfy three criteria. First, there must be a high degree of homogeneity *within* each group; second, there must be a high degree of heterogeneity *across* groups; and third, there must be a small number of significantly sized groups.

Consider how the three criteria characterise the polarisation of different societies. Figure 22.1 reproduces three examples from Esteban and Ray's original paper. In 22.1(a), a society is split into 10 equally sized groups by income bracket. In 22.1(b), we see a society split into two equal-sized groups. In each case, there is essentially the same amount of homogeneity *within* each group. However, in 22.1(a) there is a smoother gradation between the poor and the rich, with a number of people in intermediary classes. This yields less heterogeneity *across* groups than in 22.1(b). Finally, in 22.1(b) the society is just two big groups. These observations fit our intuition that the society of 22.1(b) is *more polarised* than the society of 22.1(a). This also shows that the concept of polarisation is different from the concept of *inequality*. As Esteban and Ray noted, the society shown in 22.1(b) is more polarised than the society in 22.1(a), but under any reasonable measure of inequality, the society of 22.1(b) is *more equal* than that of 22.1(a). Now compare 22.1(c) with 22.1(b). The size of the two groups is unchanged, but in 22.1(c) the first group has become much poorer and the second group much richer, so inequality has increased as we move from 22.1(b) to 22.1(c). Intuitively, polarisation has also increased from 22.1(b) to 22.1(c). This shows that while an increase in polarisation can go hand in hand with a *decrease* in inequality, it may also be the case that an increase in polarisation *increases* inequality. The two concepts are distinct.

These observations about polarisation generalise beyond income distributions. The plots in Figure 22.1 can be reinterpreted as referring to the distribution of political beliefs. This interpretation invokes what is known as the spatial theory of voting. According to this theory, people's political stance corresponds to a point in an "issue space". The simplest model treats the issue space as one-dimensional, with liberal on the left and conservative on the

**Figure 22.1: Comparisons between three different societies as an intuition pump for a measure of polarisation**




Source: adapted from Esteban and Ray (1994). Redrawn based on Figures 1A, 1B, and 2B.

right. More complex models increase the number of dimensions. In two dimensions, we might distinguish between economic policy (liberal or conservative) on one dimension and social policy (liberal or conservative) on the other. Increasing the number of dimensions increases the number of distributions, which might count as a form of polarisation. This turns out to be quite a complicated topic: Bramson *et al.* (2017) identify no fewer than *nine* different senses of polarisation in just the one-dimensional case.

In thinking about polarisation, we need to appreciate that polarisation, like many English nouns ending in *-tion*, exhibits what is known as process-product ambiguity. A word like "observation" can refer to the *process* of looking out the window – the act of observing – and the *product* of what that process yields – a person's spotting of a bird. It is important to attend to this distinction because mistakes can arise from conflating the two different meanings: things that hold for the process interpretation may not hold for the product interpretation, and *vice versa*. For example, some have suggested that a person's *perception* – understood as a process – should be treated as an unquestionable fact because, as a first-person subjective experience, how can

anyone be in a position to dispute another's experience? Yet it does not follow from this that the *perception* – understood as a product – is necessarily veracious. This distinction enables us to tell someone, "You didn't actually see what you think you saw", without denying that the person *thought* they saw what they claimed to have seen. Given this process-product ambiguity, when we speak of polarisation we need to distinguish between a process that divides people and the outcome of people being divided. This yields three questions to consider: first, are societies more polarised today than before (the product interpretation)? Second, what causes polarisation (the process interpretation)? Finally, is the type of polarisation (product and process interpretations) produced undesirable?

This last question matters because not all types of polarisation are bad. Strong differences in taste are exactly what one would expect to find in a diverse society. Strong opinions motivate people to follow their interests, to specialise in chemistry or literature, opera or rock-and-roll. Strong differences in values can prompt reflection about the kind of society we want to live in and why, helping us avoid complacency by simply continuing with current practice. But strong differences in values can lead to conflict if people feel unable to compromise on those issues that matter most to them. The overall point is that polarisation, in and of itself, is not *necessarily* bad. Polarisation may function as an engine driving a diverse and vibrant society, or it can lead to endless conflict and political gridlock.

Let us now turn to the question of whether polarisation has been increasing over the past few decades. In particular, consider whether *political* polarisation been increasing. Political differences tend to be correlated with other kinds of differences due to the close relationship between identity and politics, so this is not an unreasonable place to begin. Anecdotal evidence suggests political polarisation has increased. The 2016 US presidential election, with two competing visions offered by Clinton and Trump, elicited sharp and negative reactions from both camps when they thought of the other side winning. In the UK, the 2016 referendum on whether the UK should leave the EU similarly inflamed emotions, with a narrow margin of victory for the Leave campaign. The subsequent 2017 UK general election likewise saw no clear overall winner in Parliament, with Theresa May having to strike a deal with Northern Ireland's Democratic Unionist Party to maintain power. In both cases, the nearly equal split of the vote masks the deep ideological divide on the underlying issues. In France in 2017, Emmanuel Macron, an independent centrist, won a resounding 66–33 victory over Marine Le Pen. But the fact that Le Pen was in the final election at all, surprised people, just as when her father ran against Chirac in the final election in 2002. In Germany, the Alternative für Deutschland won over 90 seats in the 2017 election, putting a far-right, anti-Islam party – which has been accused of employing rhetoric with Nazi overtones – in parliament for the first time since World War II.

Obtaining precise measurements of polarisation for cross-country comparisons is difficult because of variations across political systems. If we restrict

attention to the US, Hare and Poole (2014) used a measure known as DW-NOMINATE to show that the polarisation of the two major political parties in 2013 was at its highest level since the end of the Civil War. The Hare and Poole method of gauging polarisation has two advantages. First, it measures the ideological position of legislators on the liberal–conservative spectrum using their entire roll-call voting record.[1] Second, it uses overlapping membership to provide a way of comparing the relative ideological positions of individuals who didn't serve in government at the same time.

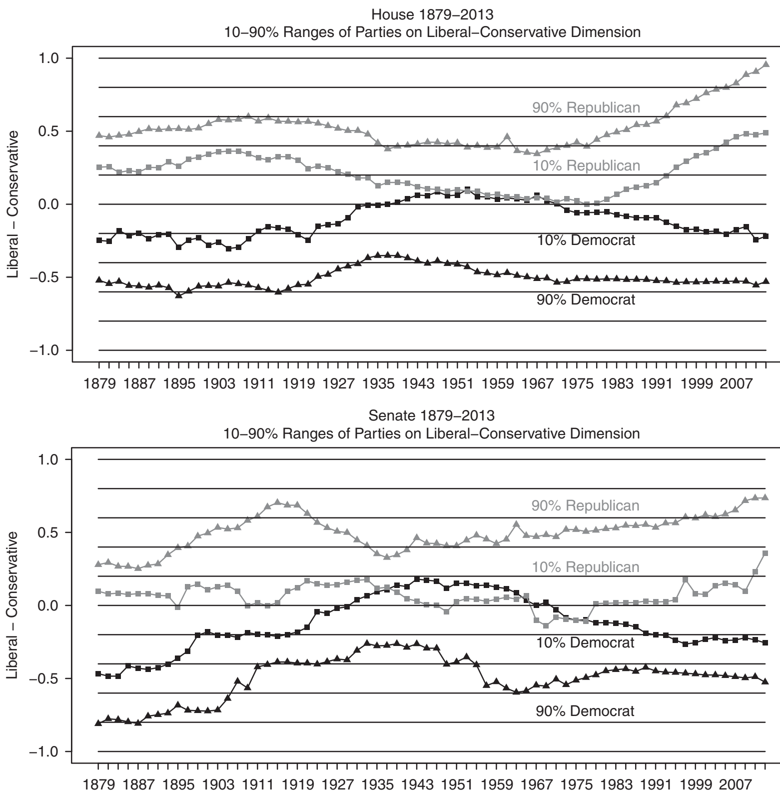
Figure 22.2 illustrates the variation in polarisation for both the US House of Representatives and the Senate from 1879 to 2013. The vertical axis represents political orientation using a one-dimensional model, where -1.0 corresponds to highly liberal and 1.0 to highly conservative. The figure shows time-series data for how the 10th and 90th percentiles of both the Republican and Democratic parties have shifted over time. That is, the 10th percentile for the Republican party tracks the point on the political spectrum such that 10% of elected Republicans are more *liberal* than that, with the 10th percentile for the Democratic party tracking the point where 10% of elected Democrats are more *conservative* than that. In the Senate, there was considerable overlap for the 40-year period from 1935 to 1975, with a number of Democrats being more conservative than Republicans and a number of Republicans being more liberal than Democrats. In the House, the period of overlap was roughly similar, although not as great. The important thing to note is how from 1975 onwards, there has been a steady divergence of the two parties.

For the UK, evidence of increasing polarisation also exists. Jonathan Wheatley examined the responses of people living in England to survey questions[2] in 2015 and 2017 (i.e., right before two general elections) and from their responses constructed a map of the ideological dimensions associated with each party's voters. Unlike the American context, in the UK it was necessary to add a second dimension to capture the full variety of individual attitudes. One dimension was familiar, measuring the extent to which a voter was economically liberal or conservative. The second dimension, rather than being indicative of whether a person was socially liberal or conservative, was instead characterised by Wheatley (2015) as a "cosmopolitan-communitarian" dimension.

What Wheatley found was striking. Figure 22.3 illustrates the ideological drift that occurred between 2015 and 2017 for all of the major political parties in the UK. What we see is a clear polarisation: the Greens, Labour, and the Liberal Democrats became more cosmopolitan and economically liberal, while UKIP and the Conservatives became more communitarian and economically conservative. Most notably, areas of overlap between the Liberal Democrats and Conservatives have disappeared, and the close proximity between some Labour voters and some Conservative voters has likewise vanished.

Although Wheatley's study only provides evidence of increased polarisation in the UK over a two-year period, other evidence suggests this is part of

**Figure 22.2: Ideological polarisation within both houses of the US Congress, 1879–2013**

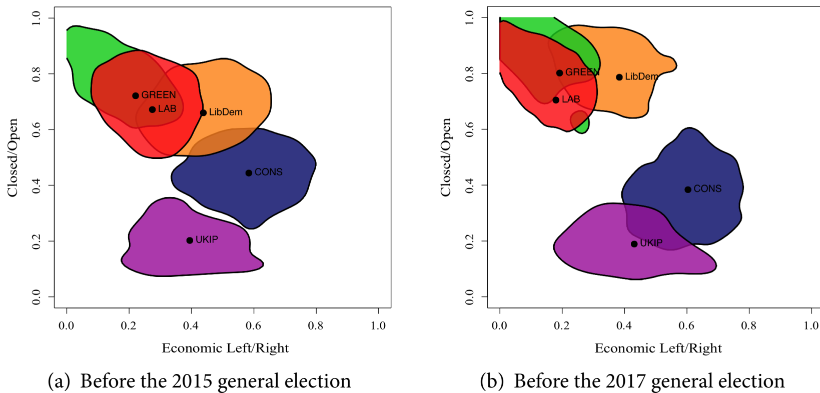



Source: Figure 1 in Hare and Poole (2014). Reproduced with permission from the University of Chicago Press.

a larger trend. Figure 22.4 shows the regional outcomes for two different referendums on whether the UK should remain part of the EC (as the EU was known prior to 1993). In 1975, the only parts of the UK to vote in favour of leaving the EC were the Outer Hebrides and the Shetland Islands, which perhaps makes sense because you already could not get much further away from the EU – physically – than those two parts of the UK. By 2016, opinion had shifted considerably towards the other end of the spectrum, with most of middle England, the south-west and part of Northern Ireland voting to leave.

These studies, although restricted to the US and the UK, are consistent with the impression that society is more polarised now than in the past. If we accept that, the next question to ask is *why*? What are the factors driving people apart?

**Figure 22.3: Ideological grouping of UK political parties before the 2015 and 2017 general elections**




(a) Before the 2015 general election (b) Before the 2017 general election

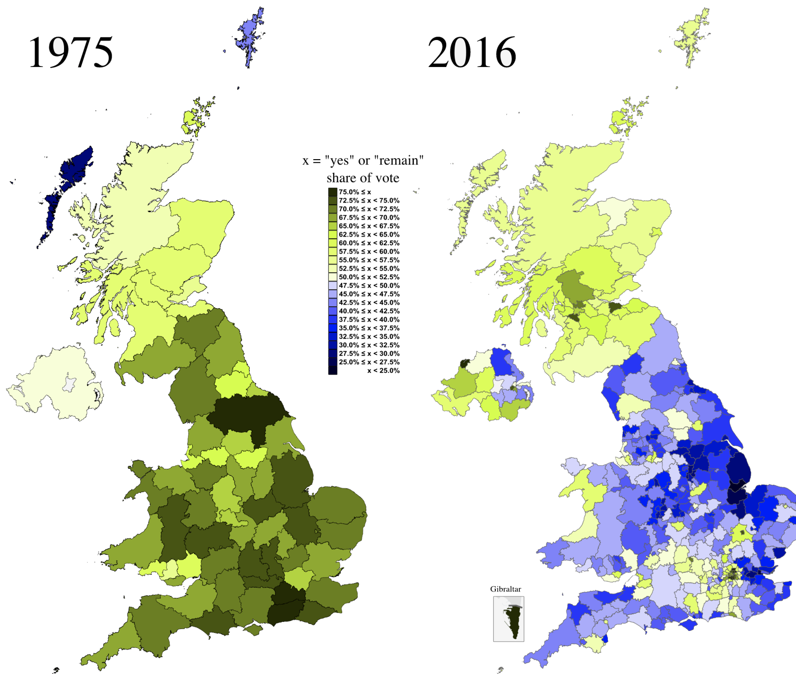
Source: Diagrams 1 and 2 in Wheatley (2017). Available under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Three possible causes of polarisation come to mind. The first is economic: since the 1980s, globalisation has become an increasingly disruptive force, displacing entire industries in the US, the UK, and elsewhere. As peoples' income in real terms stagnates or declines, they become increasingly self-interested and local in their outlook, seeing the world as a zero-sum game. This resonates with the themes covered in Part I. The second is informational: echo chambers, customised news feeds, and filter bubbles reduce the diversity of information to which individuals are exposed. This resonates with the themes covered in Part II. The third is cultural: the social identities of people have become increasingly salient, perhaps as a result of the two previous causes. And these social identities call attention to the dimensions along which people differ, rather than points of commonality.

Of these three factors, there is good evidence to support the claim that when it comes to middle-income persons living in the US, the UK, and other parts of the West, they have lost out from globalisation. Evidence regarding the effect of filter bubbles and social media is mixed, as we will see below. The last factor, of how our social identity influences our behaviour in modern tribes, shall occupy our attention for the remaining chapters in this part.

In 2012, the World Bank published a report on global inequality, examining who the winners and losers were, under globalisation. The primary beneficiaries were the very rich and the middle classes in developing economies. Figure 22.5 shows the real increase in income (measured in 2005 international dollars) versus the percentile of global income between 1988 and 2008. The largest increases occurred at the median: an 80% increase in real terms with increases of over 70% for points nearby. The report noted (emphasis added):

**Figure 22.4: Regional outcomes for the two UK referendums on whether to remain part of the EC (1975) or the EU (2016)**



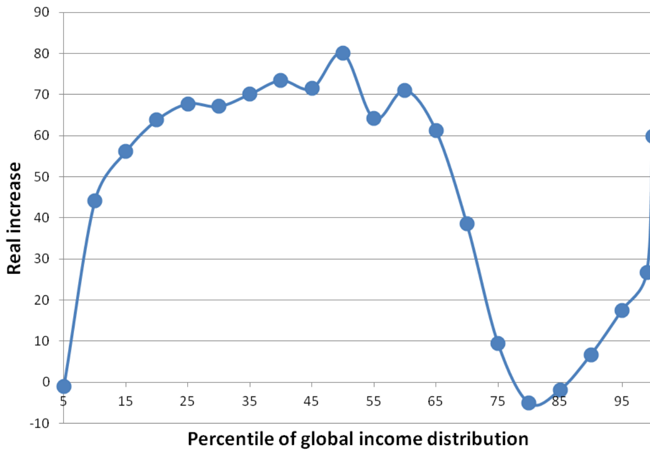

Source: Wikimedia Commons (2016). Available under a Creative Commons Attribution-Share Alike 4.0 International Licence (CC BY-SA 4.0) (https://en.m.wikipedia.org/wiki/File:United_Kingdom_European_Communities_membership_referendum,_1975_compared_to_United_Kingdom_European_Union_membership_referendum,_2016.svg).

> The most interesting developments, though, happened among the top quartile: the top 1%, and somewhat less so the top 5%, gained significantly, while the next 20% either gained very little or faced stagnant real incomes. *This created polarization among the richest quartile of world population,* allowing the top 1% to pull ahead of the other rich and to reaffirm in fact – and even more so in public perception – its preponderant role as winners of globalization. (Milanovic 2012, p. 14)

In other words, for people in the middle to upper-middle classes in developed economies, those among the richest top 15%–17% of the global population, global income distribution was *more inequal* in 2008 than 20 years prior.

What about informational contributors to growing polarisation? Sunstein observed in *#Republic* that the ability to customise online news so that each

**Figure 22.5: Change in real income between 1988 and 2008 at various percentiles of global income distribution**




Source: Figure 4 in Milanovic (2012). Available under Creative Commons License CC BY 3.0 IGO DEED.

person receives a "Daily You", displaying only that news which interests you from those sources that appeal to you, threatens to transform a well-informed citizenry, acquainted with the issues of the day, into a partisan, polarised populace where each individual is largely unaware of any perspective other than their own. How concerned should we be about that possibility?

Back in 2007, online news viewing was in its infancy. According to Baum and Groeling (2008, p. 347), "the total volume of traffic to political web sites in May 2007 was about 9 million unique viewers (Wheaton 2007), about the same as the typical audience for a single broadcast of *ABC World News Tonight*." Yet even then, worries existed that online sites were starting to show a bias towards certain kinds of news stories. Baum and Groeling analysed 1,782 AP and Reuters news stories between 24 July and 14 November 2006, along with stories from the wire services' "top news" webpages and the websites of Fox News, Free Republic, and Daily Kos. They found that concerns about one-sided coverage were warranted, as "Daily Kos on the left and Free Republic and Fox News on the right demonstrate clear and strong preferences for news stories that benefit the party most closely associated with their own ideological orientations" (Baum and Groeling 2008, p. 359). In short, online news sites were acting like a filter, selecting content based on certain ideological leanings. Recent developments with generative AI using large-language models, like ChatGPT, will only exacerbate this problem.[3]

In contrast, a few years later (Gentzkow and Shapiro 2011, p. 1801) reported "no evidence that the Internet is become more segregated over time" regard-

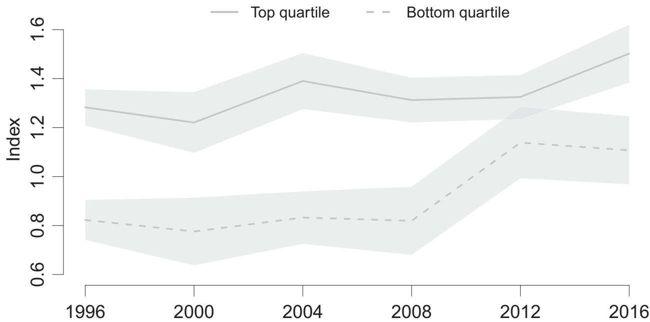
ing online news consumption. How do we account for the discrepancy between the two studies? In part, it's down to the use of different data sets and a different item of measurement: what Baum and Groeling (2008) looked at was whether *websites* displayed an ideological orientation. What truly matters, though, for the process of polarisation is the consumption of news by individuals, and this is what Gentzkow and Shapiro measured. Essentially, the reason why they found that the internet might be less effective at driving polarisation, even though many websites have an ideological orientation, is that people were varied consumers: they visited a number of different websites of different orientations, and that served to neutralise the overall effect. That said, it's worth noting that this paper was published in 2011, using data from 2004 to 2009. The world has changed a lot since then. Breitbart News wasn't launched until 2005 and didn't become a major player in the online news media until after the period of this study. For example, Robert Mercer donated $11 million to Breitbart in 2011, and Steven Bannon didn't take over as executive chairman of Breitbart until 2012. It's also worth noting that the Gentzkow and Shapiro study shows what might now strike us as a quaint concern for the importance of facts and a failure to anticipate what advances in AI might make possible:

> It is true that the Internet allows consumers to *filter* news relatively freely, but it has not changed the fact that *reporting* or *writing* stories that are tailored to a particular point of view is costly. There is no computer program that can take a story written with liberal slant as input, and output an account of the same facts written with conservative slant. (Gentzkow and Shapiro 2011, p. 1831–32)

Whereas the point about tailor-made news being costly *was* true back in 2011, that is no longer the case: AI can easily generate news on demand. And ChatGPT can easily generate new ideological perspectives on pre-existing text. For example, you might try asking it to rewrite the opening of the *Communist Manifesto* from the point of view of a free-market economist.[4] In addition, we also know that one way in which the internet polarises people is not through writing factual stories from a different point of view but, rather, through the much more effective method of "fake news", a.k.a., "making shit up."

Fast-forward to the present age and things have changed in two important ways, which make inventions like ChatGPT even more problematic. First, the amount of news that people consume online has greatly increased. In a blog post published by the Brookings Institute, Bleiberg and West (2015) wrote, "Many Americans get a significant portion of their news from Facebook and in effect the social network is the largest news platform in the U.S." Second, there has been a shift in how much news people see from certain perspectives. Bleiberg and West found that the Facebook News feed algorithm, which took a user's history of clicking on past stories into account when ranking news,

**Figure 22.6: Time-series plots of polarisation by predicted internet use**




Source: Figure 4 in Boxell et al. (2017). Reproduced with permission from the Proceedings of the National Academy of Sciences.

reduced the amount of news that people saw from the other side of the political divide by 5% for conservatives and 8% for liberals. When the disposition of users to click on links was taken into account, they found that exposure to cross-party news decreased by 17% for conservatives and 6% for liberals.

Yet there is reason to think that although these differences in click behaviour and news consumption exist, this is primarily a *symptom* of polarisation rather than a *driver* of polarisation. A study by Boxell *et al.* (2017) found that the largest changes in polarisation among US adults occurred amongst those over the age of 65, who were also the *least* likely to use the internet. A close examination of the data suggests that the effect of the internet and social media on polarisation is, at best, modest. Figure 22.6 plots the trend in polarisation by predicted internet use. Although the top quartile of internet users have a higher polarisation index than those in the bottom quartile, there is relatively little change in their polarisation index between 1996 (when little online news was available) and 2016 (when a lot was available). Furthermore, a much greater increase in polarisation occurred amongst the bottom quartile of internet users. The upshot seems to be this: whatever is driving the growth in polarisation, the internet seems to be a modest contributor.

To summarise: we have shown that there is good evidence to believe that polarisation has increased considerably in recent years. We have explored some of the possible drivers of this growth and seen that both economic and informational aspects have played a part, although economics seems to have been a much more powerful contributor than informational elements. What remains to be explored is how people's self-understanding of their *social identities* may be implicated in this polarisation trend.

It is important to keep in mind the conceptual distinction between a growth in the importance of one's *social identity* and a growth in polarisation. Or, to put the point slightly differently, we need to distinguish between *polarisation*

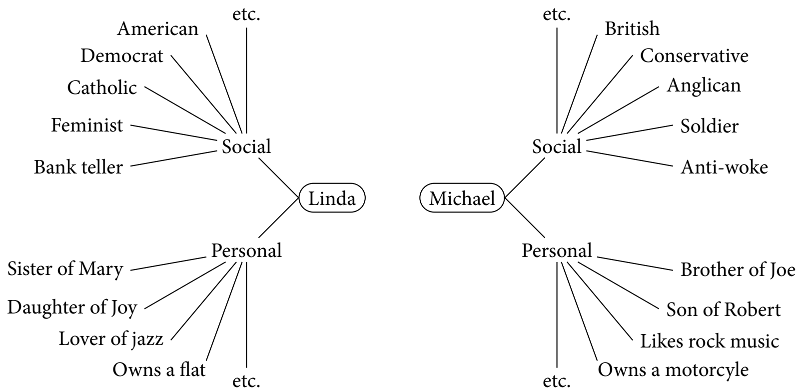
and *tribalism*. People can self-identify with a modern tribe and be highly *tribal*, even if there isn't a great deal of *polarisation*. Think of two different Christian sects whose doctrines differ by just a small amount. Although there would be little polarisation between the two groups, in terms of doctrine or beliefs, there could be a high degree of tribalism. Increased polarisation may contribute to increased tribalism, but we need to realise that polarisation is not a necessary condition for tribalism.

# 23. Social identity, in-group bias, and norms

One important part of contemporary society's division into modern tribes is how this affects our understanding of the identity of ourselves and others.[1] Social identity theory (see Tajfel 1972; Turner 1982) examines how our self-descriptions in terms of social categories, along with the normative requirements and subsequent evaluations generated by those categories, contribute to this self-understanding. Figure 23.1, based upon a diagram from Hogg and Abrams (1988), illustrates the multiple layers at play in the social identity approach. A person's self-identity features a number of both personal and social descriptions. The personal descriptions involve concrete relations with specific people (such as siblings, friends, or colleagues) or particular objects. The social descriptions, in contrast, relate the individual to a number of social *categories*, which are formally recognised or institutionalised to varying degrees. For example, in Figure 23.1, Linda's identity as an American has rigid membership criteria established by the state, but the criteria for whether she is a Democrat are less rigidly defined. A person can self-identify as a Democrat without being officially registered as such on the voter rolls. In contrast to both of these categories, what it means to identify as a feminist is even less sharply defined because there are a variety of different theoretical views, all of which can claim the term "feminist". Most of the time, a membership relation holds between the person and a specific social category, as with Michael's being a soldier or a Catholic, but sometimes a person is defined more in terms of their *opposition* to a social category, as with Michael's identity as being anti-woke (Figure 23.1).

A person's social identity attributes a number of categories, simultaneously, but only some of these categories are salient to a person at any one time. Which categories are salient, and thus influential in how a person interacts with the world, depends on environmental and individual factors.[2] In addition, membership in a tribe often comes with obligations or expectations. The obligations may concern how one should behave generally or how one should behave towards members or non-members of the tribe.[3] The expectations we have about people belonging to social categories may be based on either experience and evidence or stereotypes and, often, both. Yet there is no guarantee that these obligations and expectations, aggregated across all of the tribes a person belongs to, will be simultaneously satisfiable, much less coherent. Continuing the example from Figure 23.1, Linda's Catholicism exists in tension with her feminism, given how deeply entrenched patriarchal values are in the Roman Catholic church. If Linda really believes in a strict version of Catholicism, this belief exists in tension with the value of freedom of religion that is deeply

**Figure 23.1: The multiple layers of personal identity**




Source: author; loosely based on Figure 2.1 from Hogg and Abrams (1988).

embedded in her identity as an American because it allows other people to be *wrong* – often radically so – which would put their souls in danger.[4] Considerable conceptual gymnastics are required to resolve these tensions; sometimes it is easier to simply ignore them, hoping that they do not manifest in an explicit problem requiring resolution.

Not only does every person belong to multiple tribes but the list of tribes a person belongs to changes over the course of their life. Although relatively few people change their citizenship or religion, some do. More frequently, people change the organisations they belong to, their career, their church, and so on. And while we often think of people choosing to belong to groups because of various beliefs or the values they hold, the influence goes in the other direction as well: a person adopts new beliefs and values as a result of their membership of certain groups – beliefs and values they would not have adopted otherwise. A person's social identity thus plays an important *constructive* role in establishing a person's overall identity (Parekh 2008).

Finally, the set of obligations and expectations a person has to negotiate as a result of their membership of multiple tribes may change even if there has been no change in the *labels* that feature in a person's self-description or in the beliefs or values of that person. This can occur because of gradual *drift* in the collective tribal identity over time. Consider the following illustration, posed in the form of a trivia question: which US President created the Environmental Protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA); passed the Clean Air Act and the National Environmental Policy Act (which required many federal projects to submit environmental impact reports); put in place the Philadelphia Plan (the first major federal affirmative action programme); and endorsed the Equal Rights Amendment (ERA)

after the ERA passed both houses of Congress? Answer: Richard M. Nixon. If the answer surprised you – *surely it must be a Democrat!* – that is because the Republican tribal identity has moved to the right of the political spectrum over the past few decades. It is difficult to reconcile the Republican party that supported Nixon's legislative agenda with the Republican party of 2024. Given this, a person who described herself as a Republican during the Nixon period and continued to include that social category as part of her identity, would now find herself having to negotiate a world in which people behaved differently towards her as a result of changes in the public understanding of the "Republican" label. In addition, there is the question of how she would respond to these changes in the character of the Republican party. Would she feel marginalised as the political centre of the group became more extreme over time? Or would her beliefs and values change in order to track the median beliefs of the group?

All of this highlights the normative and behavioural complexities raised by our social identities in a world where all of us belong to multiple tribes. Aside from the inner conflict this generates for us, there is also the outer conflict it generates because different groups have conflicting, or only partially aligned, interests and goals they seek to realise. But are there any other phenomena we should be concerned about which arise simply in virtue of the fact that people see themselves as members of a *group*?

In a classic result from social psychology, Tajfel *et al.* (1971) showed that simply putting people into *groups* sufficed for inducing bias towards one's fellow group members, known as the *in-group*. What is striking about the phenomenon of in-group bias is that it can be created in people simply by virtue of the fact that they see themselves as *belonging to a group*, even when there is no salient collective identity that binds the group together. It seems that our disposition towards tribal thinking has a hair trigger indeed.

In one experiment, subjects were shown a series of photographs containing a number of dots and were asked to estimate how many dots were displayed. Once their estimates were collected, the subjects were divided into groups, ostensibly on the basis of whether they tended to overestimate or underestimate the number of dots on screen. In truth, subjects were put into groups at random. A second experiment altered the initial stage by showing subjects a series of two abstract paintings, without identifying the artists, and asking people to state which of the two paintings they preferred. Subjects were then divided into groups on the alleged basis of whether they tended to prefer Klee or Kandinsky (the featured artists). As before, group allocation was actually done at random.

Once subjects were assigned to groups, they were seated in isolated cubicles and given a booklet containing a number of two-person resource allocation problems. The allocation problems were stated using a matrix of numbers, with each column representing a possible outcome and the payoffs to the two players given by the numbers in the corresponding rows. Each row was labelled: "These are rewards and penalties for member $N$ of your group"

or "of the other group", with $N$ being the code number of the person to ensure anonymity. The allocation problems fell into three types: *in-group* problems, where both recipients were from the same group as the subject; *out-group* problems, where both recipients were from the other group; and *inter-group* problems, where one recipient was from the in-group member and the other from the out-group. Subjects had to indicate which outcome they wanted by marking the appropriate column.

What Tajfel and his colleagues found was striking: when choosing allocations for inter-group problems, subjects showed a clear tendency to *favour* fellow in-group members at the expense of out-group members. This in-group bias was present despite the lack of any apparent reason for showing such favouritism: why would the mere fact that a subject was placed among fellow over- (or under-) estimators of dots prompt an aversion to fair distributions? Other experiments (see, for example, Nydegger and Owen 1974) found that in perfectly symmetric situations, people preferred to divide a resource fairly. Here, in a situation where people are essentially symmetric and the only difference is which group they belong to – this alone suffices for breaking the symmetry. Furthermore, it's important to remember that, in making these allocations, the subject's choices concerned *other* people, not themselves. And, since there was no possibility of communication with fellow group members after the allocation problems were revealed, there is no reason for subjects to think that any in-group favouritism, on their behalf, would be reciprocated. Rather, there seemed to be something merely in the act of *seeing oneself as belonging to a group* that triggered in-group bias. These results have been replicated in a number of other settings. If you harbour a worry about whether these results were truly *minimal* enough, in that the stated reason for group membership did refer to a criterion of similarity (i.e., similar tendencies to over/underestimate dots, or preference for a particular abstract artist), in a follow-up experiment, Billig and Tajfel (1973) found that in-group bias still occurred when subjects were put into groups *and told that group membership was assigned at random*.

If the mere awareness of belonging to a group – even a group with no meaningful identity – can cause people to display favouritism towards fellow group members, more extreme kinds of behaviours are easily induced when group membership is based on things people actually care about. This, of course, is no great surprise: groups of people have been killing each other in warfare throughout all of human history. What is surprising is how easily group membership, combined with a little competition, can give rise to hostilities when previously no animosity existed between individuals.

In order to understand this phenomenon, we need to realise that real-world group membership – unlike the minimal group experiments discussed previously – quickly becomes associated with a set of values and norms that regulates the expression and enforcement of those values. Interactions between people belonging to different groups then frequently cease to be seen as two

people interacting *qua* individuals and are instead seen as two people interacting *qua* group representatives, with the potential conflict of values that entails. One highly influential theory of inter-group relations, known as realistic conflict theory, advances the following hypotheses:

> When a group forms it delineates itself (ingroup) from an outgroup. This categorical distinction then comes to embody value-laden content. Ingroup norms develop from interpersonal relationships within the group which define the range and content of acceptable ingroup values, and the rewards or sanctions associated with adhering to these norms. Stereotypes are then applied to outgroups, the content of which depends on the actual or perceived relations between the groups in question. Specifically, if the groups are seen as being in competition, such that something which is good for one will be bad for the other, the stereotype of the outgroup is likely to be negative and derogatory. (Hogg and Abrams 1988, p. 43)

This theory of inter-group relations was tested in a famous series of experiments conducted by Muzafer Sherif in 1949, 1953, and 1954 (see Sherif 1962, 1966).[5] In the 1949 experiment, Sherif and his collaborators arranged for 24 boys, between 11 and 12 years of age, to attend a summer camp in northern Connecticut. The boys were carefully selected so as to minimise any naturally salient attributes that could induce divisions within the group. All the boys were white Protestants from "settled American families of the lower-middle-class income group in the New Haven area" (Sherif and Sherif 1953, p. 238). They all had similar educational backgrounds and all were classified as more or less "normal". The most important point is that none of them were considered to have any behavioural problems.

The experiment had three stages. In the first stage, the boys – all strangers to one another – lived in a single large communal cabin where they were allowed to develop friendships. In the second stage, the boys were split into two groups of equal size and moved into separate bunkhouses. The groups were kept separate as much as possible, with no common activities. Over the five days of the second stage, each of the groups began to form a collective identity. They adopted names for their groups (the "Red Devils" and the "Bull Dogs", respectively), made T-shirts with their group mascots, and built private hide-outs. Different behavioural norms became established in each group, ranging from how they punished social infractions by members to how they made lanyards or cleaned up trash. When the boys were asked an open-ended question about who their best friend was, with the wording indicating that they could name anyone in the *camp*, almost everyone answered by naming people within their *group*.

The final stage of the experiment involved putting the two groups into competitive situations where the aims of each group were frustrated. This began

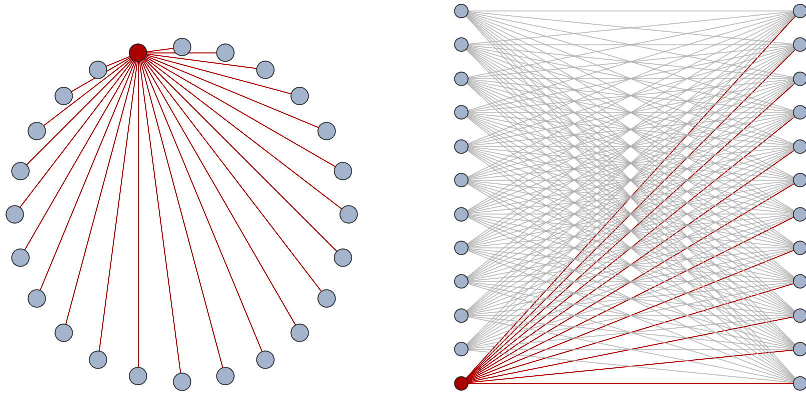
with three days of sporting competitions, with prizes given to the winners. Although the boys started with a show of good sportsmanship, their attitudes changed over the course of the competitions. A cheer which began, "2–4–6–8, who do we appreciate?" morphed into "2–4–6–8, who do we appreci-*hate*?" over the course of a match. When it became clear that the Bull Dogs were going to win the overall competition, the Red Devils were calling the Bull Dogs "cheats" and "dirty players", blaming their loss on unfair play. Other situations were engineered to generate further conflict between the groups. This led to increasingly hostile interactions, bordering on outright violence. This experiment wouldn't receive ethics approval today.

Belonging to a tribe with attitudes of group loyalty increases the likelihood of conflict occurring. Suppose, following the Sherif experiment, we have a population of 24 people, initially without any salient group identities. In this case, there are "only" 23 interactions that might get someone into a fight: one interaction for each of the other people in the population, as Figure 23.2(a) shows. Now suppose the population is split into two groups of 12, the Bull Dogs and the Red Devils, and that a person *never* gets into a fight with a fellow group member. Suppose, though, that any time a fellow group member gets into a fight, all their friends join in. Then the number of interactions that could possibly get a person into a fight increases to 144: the 12 interactions that person has with the other group, plus the twelve interactions for every one of 11 fellow group members. This is illustrated in Figure 23.2(b). The most important thing to note is that this generalises: every social group one includes as part of one's overall identity adds a whole set of new pathways for experiencing approbation or admonishment.

Sherif's work also highlights the role played by stereotypes in inter-group dynamics. In the absence of salient social categories, people must interpret their interactions with someone in *individual* terms. If I'm playing a one-on-one game of basketball and my opponent shoves me a bit too hard while trying to make a shot, I might think, "*Did he mean to do that?*" If I'm playing five-on-a-side, and I get shoved by two in quick succession, I might think,"*These guys play really rough!*" From that hasty generalisation, a stereotype emerges. The stereotype then provides a frame for interpreting future interactions for the rest of the game. As the size of the other group increases, the harder it becomes to ground one's interactions on an informed, nuanced understanding of the other person's unique character traits.

Stereotype theory has developed considerably since the early work of Allport (1954) on the nature of prejudice. Allport's original theory primarily focused on negative stereotypes built around the in-group/out-group distinction reflected in Sherif's experiment. More recent work, such as the stereotype content model of Fiske *et al.* (2002), treats stereotypes as having a more complex structure. According to this model, stereotypes vary along two dimensions: warmth and competence. This yields the $2 \times 2$ classification of stereotypes as shown in Figure 23.3. Stereotypes applying to the in-group or close allies

**Figure 23.2: Adding group structure introduces a number of new pathways along which conflict might emerge**



(a) The 23 interactions that might get a person into a fight when no group structure is present

(b) The 144 interactions that might get a person into a fight when a group structure is present. The interactions that the person is directly involved in are highlighted; the greyed-out interactions indicate the interactions of fellow group members.

Source: author.

tend, not surprisingly, to be high in both warmth and competence.[6] Stereotypes applying to out-groups can vary considerably depending on whether we see ourselves in competition as well as on perceptions of competence.

There's one last piece of the puzzle: the nature of norms. Group identity is often associated with norms governing members' behaviour. Bicchieri's theory of norms, which we've encountered before, distinguishes between *descriptive* norms and *social* norms. A descriptive norm "is a pattern of behavior such that individuals prefer to conform to it on condition that they believe that most people in their reference network conform to it" (Bicchieri 2017, p. 19). People won't follow a descriptive norm regardless of what others do, it depends on what other people do. If *no one* was ever fashionably late to a party, you'd probably always be on time, but since most people are fashionably late, you'll probably be fashionably late too. Second, the "other people" who matter are determined by the person's "reference network", and this may vary from person to person. As Bicchieri (2017, p. 19) notes, "A young woman in Philadelphia wearing very high heels will probably not care what other women do in India, or even New Orleans." The young woman's reference network is restricted to "the 'fashionable' crowd in her town, those who she is likely to meet" or perhaps certain celebrities or fashion icons.

**Figure 23.3: A $2 \times 2$ categorisation of stereotypes**

| | | Competence | |
|---|---|---|---|
| | | Low | High |
| Warmth | High | Paternalistic prejudice<br>Low status, not competitive<br>Pity, sympathy<br>(e.g., elderly people, disabled people, housewives) | Admiration<br>High status, not competitive<br>Pride, admiration<br>(e.g., the in-group, close allies) |
| | Low | Contemptuous prejudice<br>Low status, competitive<br>Contempt, disgust, anger<br>(e.g., welfare recipients, poor people) | Envious prejudice<br>High status, competitive<br>Envy, jealousy<br>(e.g., rich people, feminists) |

Source: Adapted from Fiske *et al.* (2002).

Recall Bicchieri's definition of a social norm, mentioned in Chapter 13:

> A social norm is a rule of behavior such that individuals prefer to conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation). (Bicchieri 2017, p. 35)

Whereas a descriptive norm is a "pattern of behaviour", a social norm is a "rule of behaviour". That's an important difference: *rules* are normative in a way *patterns* are not. There's no rule saying you *must* be fashionably late to a party. In addition, with a social norm people conform to the rule because other people believe they *should*, where the "should" is a normative statement and not merely a prudential one. Social norms are rules people follow because it is the way things are done and *because other people expect you to conform*; it's not simply a matter of it naturally being in your self-interest.

Putting these elements together, it's clear how polarisation can become dangerous. The tendency to treat out-group members differently seems hardwired into our psychology. Because our concept of identity includes our *social* identity as an important constitutive part of *who we are*, we cannot avoid having our group membership – and, hence, the group membership of others – constantly made salient. The variation in social norms across groups provides multiple ways two groups can be perceived as different. Because social norms set expectations about how one *ought to behave*, differences in social norms

between groups mean that members of two different groups may perceive the other as not behaving as how one ought to behave. Increased group polarisation increases the degree of perceived social distance between groups. This process of "othering" plays into our hard-wired tendency to treat out-group members differently. In extreme cases, out-group members can be dehumanised, enabling the commitment of great atrocities. This is why, in the next chapter, we consider the psychology of modern tribes.

# 24. The psychology of modern tribes

In modern tribes, social identity theory, in-group bias, stereotypes, and social norms interact. Social identity theory tells us that we are taught from a young age to identify with a number of tribes and incorporate that into our sense of self. As we get older and develop views about politics, morality, and other matters, we can augment that self-identity by joining additional social groups that speak to those interests. The fact of in-group bias means that we probably will, from time to time, show unwarranted favouritism towards in-group members. These acts of favouritism may be explicit or implicit. Implicit acts of favouritism are difficult to identify. If I attend a party and only engage in conversation with a few people over the course of the evening, is that an accident of social dynamics or the result of the deliberate avoidance of others? In real life, it can be a bit of both.

As we saw, the minimal group paradigm experiments show that people's tendency to favour in-group members can be invoked even when group identity isn't relevant. That's one possible source of conflict. When we are on the receiving end of out-group bias, it reinforces the negative stereotypes associated with out-groups. This matters because when we encounter an out-group member we don't know, we often frame the interaction using stereotypes. When coupled with the psychological phenomenon known as "out-group homogeneity", which means people tend to view out-group members as largely undifferentiated, the negative stereotypes are readily applied. Finally, patterns of behaviour generated by these psychological processes can become encoded as social norms.

Once social norms form, group divisions and poor inter-group relations can be maintained over time, even if people dislike it. In-group members who hesitate to behave negatively towards out-group members may be sanctioned for failing to comply with the norm. In-group members who comply with a norm of discrimination towards out-group members recreate expectations of such behaviour in the future. How so? Their witnessed compliance with the norm in the *present*, reinforces *other* in-group members' belief that the empirical expectation condition is satisfied. (That is, most people in one's reference network conform to the rule.) A person's compliance with a norm also reinforces *other* peoples' beliefs that the normative expectation condition is satisfied. Why? Because when a person complies with a norm, that provides evidence that person thinks others ought to comply with the norm as well. If a person hesitated to conform with the norm, but did so after being sanctioned,

that shows they didn't think nonconformity was worth the price they had to pay.

Those are the basic facts about the psychological dynamics of modern tribes. Our tribal existence faces further complexities given social media and how much of modern social life is online. Before exploring the negative aspects, let me acknowledge the important positives. One of the real advantages of social media is how easy it is to find other people like you, with your interests. Social media, and the internet, provide a forum where people can coordinate and interact, whatever their interests. That has the potential for incredibly positive and self-affirming interactions. It is useful for disadvantaged or marginalised individuals who would find it otherwise difficult to mobilise and be heard. It also has allowed powerful social movements such as #BlackLivesMatter and #MeToo to achieve international recognition more rapidly than previously possible.

Yet one downside is that when inter-group interaction takes place via social media, there are few checks and balances in place to counteract the harmful behavioural tendencies that can be triggered when group identification becomes salient. Although many social media companies provide users with the ability to control who sees their posts, many people don't exercise this control. This allows online interactions to occur between largely anonymous persons.

There is an interesting theoretical question as to what exactly "anonymity" amounts to. The main connotations are that a person is unknown, nameless, or unidentifiable. But, as Lapidot-Lefler and Barak (2012) point out, anonymity then becomes context specific. They observe that the mere absence of a name on a publication such as a newspaper article might suffice to make someone unidentifiable. But being nameless may not suffice in the context of face-to-face communication, where identifiability can result from other personal attributes. In what follows, I shall treat "anonymity" as a property, admitting of degrees, which measures how much person $A$ knows about person $B$. To say that a person is "largely anonymous" means that there is very little known about that person but that there is some information available from which further inferences can be made.[1]

Consider, then, what happens when largely anonymous persons engage in an exchange regarding highly value-laden matters such as race relations, gender, immigration, economics, politics, and so on. When this happens, the interpretation of what is said by the anonymous participants often draws heavily on stereotypes triggered by the minimal information available. As we saw from the stereotype content model, out-group members can fall into one of three categories. The two competitive categories with low warmth involve stereotypes centred around either contemptuous prejudice (for low competence groups) or envious prejudice (for high competence groups), with attitudes of contempt, disgust, anger, envy or jealousy. With those stereotypes influencing how one interprets the inter-group exchange, it does not bode well for civil and charitable discourse. *Especially* when we recognise that the salient group identities will not be "minimal groups" of the kind Tajfel studied

but, rather, groups expressing value-laden content where those values feature prominently in a person's identity.

We've already discussed online shaming in Chapter 11, but we can now supplement that analysis with a more informed understanding of human psychology. The eagerness of people to respond critically to online statements of others is driven by the psychology of inter-group conflict: negative stereotypes influence the interpretation of out-group members, and in-group members contribute to online shaming in order to signal commitment to the in-group. (This behaviour is closely related to "virtue signalling".) Any in-group member who expresses dissent or concern about joining Two Minutes Hate[2] may find themselves sanctioned by in-group members. These enforcement policies can themselves become a norm, what the social Scientist Robert Axelrod called a *metanorm*. Axelrod (1986, p. 1100–01) provides the following vivid illustration of a metanorm in action:

> A little-lamented norm of once great strength was the practice of lynching to enforce white rule in the South. A particularly illuminating episode took place in Texas in 1930 after a black man was arrested for attacking a white woman. The mob was impatient, so they burned down the courthouse to kill the prisoner within. A witness said,
>
> > 'I heard a man right behind me remark of the fire, "Now ain't that a shame?" No sooner had the words left his mouth than someone knocked him down with a pop bottle. He was hit in the mouth and had several teeth broken.' (Cantril 1941)

Tribal interactions on social media take place using impersonal, remote forms of communication rather than personal, direct forms of communication. That might sound like an obvious truism and thus unimportant, but some have thought this partially explains why so many people behave so badly online. Suler (2004) provided the first analysis of some causes behind what he called the "online disinhibition effect". He suggested six factors that contributed to online disinhibition: anonymity, invisibility, asynchronous communication, minimisation of status and authority, dissociative imagination (that people do not fully appreciate that what they do online is part of the "real world"), and solipsistic introjection (that people conceive of their online interlocutor as a character within their own head). Later experimental work has suggested some of these factors are more important than others. Lapidot-Lefler and Barak (2012) argue that when subjects engage in pairwise interactions controlled for anonymity, invisibility, and eye contact, the presence or absence of eye contact matters most regarding the hostility of the exchange. If the online disinhibition effect exists, and tribal identities are made salient and

put into conflict, and those tribes have become more polarised over time – we have a partial explanation of the tempestuous nature of social media.[3]

There are two generally recognised mechanisms for reducing inter-group conflict. The first attempts to undo the negative stereotypes that lie at the heart of the problem. This method, due to Gordon Allport, is known as the contact hypothesis or inter-group contact theory. The contact hypothesis says that if members of two conflicting groups are able to have positive interactions with each other, under the right conditions, this reduces prejudice and mitigates stereotypes. Inter-group contact challenges the unconscious assumption of out-group homogeneity by getting a person to see out-group members as unique individuals with interests, goals, and ambitions, which may partially overlap with their own. In addition, the contact allows one to understand what the world looks like from the other viewpoint.[4]

The second method, identified by Sherif in his 1954 experiment, was for the two conflicting groups to be faced with a common threat that endangers the well-being of both groups such that only cooperation could resolve the threat. This mechanism – known as the existence of a *superordinate goal* – appeals to the self-interest of both parties, requiring cooperation to achieve it. Think of the scene from *Raiders of the Lost Ark* where Indiana Jones and a thug are choking each other during the fight in Marion's bar. When Toht says: "Shoot them. Shoot them both", Indy and the thug immediately stop choking each other and collectively fire a gun to eliminate the threat. That's a perfect example of a superordinate goal eliminating conflict.

Both the contact hypothesis and the existence of superordinate goals have been shown to be effective at reducing prejudice and inter-group conflict. However, there are several open questions about how well these experimental findings carry over into real-world situations. To begin, as noted before, there is a great difference between face-to-face contact, where each subtle aspect of human communication is visible and noticeable, and online communication. If the goal of contact is to generate a perception of "common humanity between members of the two groups" (Allport 1954, p. 281), online communication makes that difficult. That said, a number of studies have examined whether computer-mediated communication is compatible with the reduction of prejudice suggested by the contact hypothesis (see Hasler and Amichai-Hamburger 2013, for an example) with encouraging results. But the downside of these studies is that they typically use *highly controlled* online environments. When we contrast that with the wild west of unmoderated social media, different results obtain. In one study, Ruesch (2013) examined 770 groups on Facebook, each of which had over 100 members, organised around the Israel-Palestine conflict. What she found was "a highly fragmentised, polarised virtual sphere with little intergroup interaction" (Ruesch 2013, p. 22). And, perhaps not surprisingly, "Facebook groups are rather used to indicate support and opinion than to deliberate with the non-like-minded" (Ruesch 2013, p. 20). It seems that uncontrolled interaction online is more

likely to bring about homophilic self-assortment than the kind of inter-group interaction required by the contact hypothesis.

As for the efficacy of superordinate goals as a means for combating inter-group conflict, the appeal to common self-interest only works as long as the threat is truly perceived as a threat to the self-interest of both parties. Depending on how the group identity is defined and how people understand their relation to the group, this might not always lead to an accurate perception of the underlying threat of the superordinate goal. The point, here, is similar in spirit to Upton Sinclair's observation that "it is difficult to get a man to understand something, when his salary depends upon his not understanding it" (Sinclair 1995 [1934], p. 109). We might similarly say that it is difficult to get a person to understand something when that person's *identity* depends upon their not understanding it.

As an example, think of the issue of anthropogenic climate change. As a superordinate goal faced by nations around the world, and by political parties within nations, it is difficult to conceive of a greater existential threat to the human species, except perhaps nuclear war. Given that, one might expect climate change to provide a superordinate goal that would reduce conflict between political parties in the US *at least on that issue*. But the problem is that denial of climate change has become a core part of the Republican group identity. The long-term time horizon of the problem allows efforts to reduce climate change to be reframed as a zero-sum game between two political parties. The action required to avert climate change can interpreted by Republicans as an attack on their political, economic, or social values. When the constituents of group identity involve commitments that impair one's ability to engage with reality accurately, superordinate goals can lose their effectiveness at mitigating inter-group conflict.

The fact that a person's group identification can, at times, impair their ability to engage with reality – and sometimes cause them to make decisions that run counter to their self-interest – is of great importance. It is especially so because, much of the time, the criteria used to determine whether a person is a "true" member of a group is set by social norms, which are not under any single person's explicit control. When a person's group identification matters greatly to the person, they can find themselves doing things in order to prove their *bona fide* membership in the group, even doing things that they would not normally be disposed to do. We turn to the subject of *authenticity* in the next chapter.

## 25. Authenticity and the WINOs

As we've seen, identifying with groups is an important part of a person's identity. One question that we haven't yet broached concerns the *membership criteria* that determine whether some person $X$ belongs to some group $W$. Closely related to this, but conceptually distinct, is the question of what it *means* to be a $W$, from the point of view of $X$ as well as other members of $W$ and the wider society to which $W$ belongs.

The reason why the membership criteria and what it means to belong to a group are distinct can be seen by reflecting on some of the cases discussed previously. In many of the cases we've looked at, the membership criteria are trivial: a person belongs to a group because they were simply *assigned* to it. This was true for Tajfel's minimal group paradigm experiments as well as for the Sherif experiments. The Bull Dogs and Red Devils were classified as such because Sherif's team created those groups. Yet even though the membership criteria was trivial – people belonged to the group by fiat – those groups developed a distinct identity over time. What did it mean to be a Bull Dog? Among other things, it meant that they worked well as a team and had greater organisational skills than others; this was a self-understanding that each Bull Dog came to appreciate. In real life, groups like this are found all over the place: educational classes where students are grouped by ability, working groups created by a boss, and so on.

In other cases, people aren't assigned to a group but find themselves belonging to a group because some administrator, somewhere, laid down a definition of a category and people found themselves in a group because they met the definition.[1] Here, the membership criteria are whatever the administrator declared. Whether there is any deeper meaning associated with the group depends on the role that category plays in wider society and how group members come to understand it. Sometimes the administrative category becomes part of people's identity with meanings and values that go beyond the original definition. Consider the term "middle class", for example. It was coined by the Irish statistician T. H. C. Stevenson in the 1913 Report of the UK Registrar-General. Since then, the category has acquired a number of connotations regarding normalcy that people see as desirable. It's so desirable that, in a 2015 survey, 89% of Americans considered themselves to be "middle class" (Pew Research Center 2015) even though 13.5% lived in poverty (Proctor *et al.* 2016). That's already 102.5%, and we haven't said anything about the wealthy or the rich.

Ethnic terms have begun life this way too. “Hispanic” and “Latino” entered into mainstream usage in the US as a result of their inclusion in the US Census during the late 20th century, morphing over time into quasi-racial terms with cultural associations. And this happened *despite* the US Census clearly indicating that it was an ethnicity not associated with any particular race. In a 2004 survey, 400 Dominican immigrants in New York City and Providence, Rhode Island, were asked how they defined themselves racially. The question was open ended so people could answer however they wanted. In response, 27.5% said “Hispanic” and 4.1% said “Latino” as their racial self-categorisation (Itzigsohn 2004).

In contrast, other kinds of groups are fluid with both the membership criteria and the meaning of membership. Of this latter type, consider a person’s political identity. What, if anything, *must* someone be committed to if they describe themselves as liberal or conservative, Democrat or Republican, Labour or Tory? The current state of society plus recent history set general expectations about a person’s beliefs and values when someone declares their political alignment, but those expectations are defeasible. We might be surprised to hear a self-declared liberal oppose abortion or the welfare state or declare support for nuclear weapons, etc., but there are reasons – principled reasons – why a sincere liberal might endorse any one of those positions. Political alignment and party affiliations are a broad church, with room for diverse combinations of opinions.

In groups of this kind, it’s not uncommon for struggles to erupt over what exactly is required for someone to claim that they are a member of group $W$. Often the language of *authenticity* is used. A person who claims to be a $W$, who expresses a view on a matter relevant to, perhaps even central to, the $W$-identity, is criticised by other $W$s for not being an *authentic* $W$. As such, their view as an inauthentic $W$ is discredited, disregarded, delegitimated, or marginalised. The person is labelled a $W$ “in name only”, meaning that, although they *claim* to be a $W$, they aren’t a *real* $W$.

These people are the WINOs: a “(Whatever) In Name Only”. In recent years, there have been an awful lot of WINOs. In America, there are the RINOs (“Republican In Name Only”) and the DINOs (“Democrat In Name Only”). In the UK, after the divisive referendum campaign regarding membership in the EU, those Brexiteers who have argued for continued membership in the Customs Union and the single market have been called BRINOs (“Brexiteer In Name Only”). If you were a moderate Republican in America, you might have thought it was bad enough being a RINO, until you met someone wearing a t-shirt saying they were a “RINO hunter”. What does a RINO hunter do? Presumably root out the RINOs corrupting the rest of the party, with the implicit threat of violence not lost on anyone.

Although the term RINO is quite recent, with the first known use appearing in print in 1992, the phrase it abbreviates dates back much further. In an article entitled “A Strange Blunder”, published in *The National Republican* on 26 January 1875, we find the following rant against two politicians:

> Next on the list, beginning from the same end, we find Mr. William Walter Phelps, of New Jersey, and Mr. Charles Foster, of Ohio, both of whom are Republican in name only, and both of whom have proved their treachery to party principles, to party friends, and to the policy which can alone secure the success of the party.

The sentiment behind the expression hasn't changed much in over 140 years: a claim of treachery, a charge of disloyalty, and a statement of posing a threat to the success of the group. But, aside from the hostility faced by RINOs, and WINOs more generally, what further concerns are raised by charges of inauthenticity?

The first issue when a WINO is charged with inauthenticity is, who decides what it means to be an authentic $W$ and on what authority? In thinking about this, note that there are two different concepts of authenticity in play: authenticity of the individual and authenticity of the group. One reason people care about authenticity is because of our individualistic society, with the ideal that *you should be true to yourself* and resist pressure to act otherwise. This is one reason I argued for an existentialist foundation for cosmopolitanism in Chapter 4. Yet what renders an individual authentic is different from what makes an individual an authentic member of some group $W$, and so the two notions may conflict.

As we have seen, a core idea of existentialist thought is freedom of choice, wherein individual choices are, in part, attempts to realise some end that person values. This doesn't mean a person has complete control over the outcome; the world places limits, sometimes severe ones, on what a person can achieve. Yet these limits do not remove the freedom of choice, although they may restrict substantially the number of options from which one has to choose. To deny this freedom and choose contrary to one's beliefs and values, perhaps by caving in to social pressure, is to act in *bad faith*. A person may feel compelled – given their beliefs and values – to act one way, while feeling compelled – through the peer pressure applied by social norms – to act differently. In such a case, the competing demands of individual authenticity and group authenticity pull in different directions.

Suppose, for purposes of illustration, you consider yourself to be Catholic but believe abortion is not necessarily wrong and should not be prohibited, as long as certain conditions are met. In particular, suppose you do not believe life begins at the exact *moment* of conception. You might not have a firm belief on when exactly life begins between conception and birth, but your inability to answer that question does not preclude you from believing that abortion is permissible in the early stages of pregnancy. Given these beliefs, you are then confronted with the fact that the official doctrine of the Catholic Church[2] requires the excommunication of Catholics who have an abortion outside the narrow conditions that render it morally permissible. Suppose that you, then, face a situation in which you have to choose whether or not to have an abortion

under conditions that you believe to be morally permissible. If you choose contrary to your beliefs due to the threat of being excommunicated, you are being inauthentic and acting in bad faith. Yet if you act in accordance with your beliefs, you risk excommunication by the Church, thereby being declared not to be an authentic Catholic.

Although the source of the conflict between individual authenticity and group authenticity seems straightforward – a person's beliefs and values pitted against those beliefs and values required for group membership – social identity theory suggests that the conflict reoccurs deep within the person. In fact, social identity theory challenges the very idea of what it means to be "true to yourself" in the face of external pressure. Our social identities are incorporated into our overall identity in ineliminable ways. The flow of information and influence goes in both directions – from the person to the group and from the group to the person. Given this, where do I draw the line between *my* beliefs and values and the beliefs and values I acquire in virtue of my identifying with a group?

In order to answer this question, we first need to clarify how we think of our beliefs, desires, and values. One model, which I think many people employ, is what we might call the Library Model. According to the Library Model, a person's beliefs, desires, and values exist within their head, much as if they were books on shelves inside the library of the mind. When we introspect about whether or not we believe something, or have a desire, we check to see if that "book" is present in our mental library. If it is, we open the book and see what is written inside: the contents give our attitude towards the belief, desire, or value, along with the reasons (if any) for why we hold it and any relevant cross-references to other "books" (i.e., other beliefs, desires, or values) to which it relates. The absence of a book indicates the absence of any attitude towards that belief, desire, or value. This allows us to distinguish between having no opinion about an issue (e.g., "Gosh, I don't know. I've never thought about that before.") and having a definite opinion of no judgement (e.g., "I've thought about this long and hard and have come to the conclusion that the evidence does not settle the matter either way.").

The trouble is that the Library Model doesn't quite capture the complexity of our mental lives. First, it assumes that we only have attitudes on issues which we have actively considered at some point in time. As Pettit (1995) points out, many of our beliefs – and, I would suggest, many of our desires and values too – are ones that we only hold *virtually*. If I ask you how old you are, your belief that you are $N$ years old will be retrieved from memory because you have actively considered it in the past. However, if I ask you whether there are more grains of sand in a 100ml jar than people living in Bob Dylan's hometown of Hibbing, Minnesota, I am willing to bet that there will be no belief stored in memory to retrieve. Rather, you will most likely infer, given *other* things that you believe, that the answer is yes.[3] Since it sounds odd to say that you didn't *believe* that, prior to being asked, Pettit introduces the concept of a virtual belief to cover this case. The Library Model doesn't easily extend to cover those

virtual attitudes that readily follow from your current attitudes but which you haven't ever considered.

There's a second area where the Library Model proves to be inadequate. In "The Extended Mind", Andy Clark and David Chalmers argue for a view known as *active externalism*, wherein beliefs don't just exist in the head. Consider the following thought experiment. Suppose Otto is a person suffering from Alzheimer's disease who, in order to cope with his failing memory, records information in a notebook, which he always keeps with him. When Otto decides to go see an exhibition at the Museum of Modern Art, he consults his notebook, reads that the museum is located on 53rd Street, and goes there. Although Otto didn't have the location of the Museum of Modern Art stored as a belief inside his head, he was able to use the information stored in his notebook to get him where he wanted to go. Otto's notebook serves as a source of information which can be used to guide his actions, thus functioning essentially like his memory. The main difference is that it isn't inside Otto's head.

Although the original formulation of the Library Model wasn't designed to accommodate the extended mind, it can be easily modified to do so. Let the *books* in the Library Model represent the individual attitudes (beliefs, desires, or values) stored in the library of the mind. In addition, let's allow for slips of paper to appear on the shelves alongside the books. Each slip of paper represents an instruction redirecting the person to another source where the appropriate attitude can be found. A kind of interlending scheme for the library of the mind, if you like. The important point is that the slips of paper pointing at another source aren't restricted to referring to inanimate objects like Otto's notebook, they can point to other *persons* as well. Clark and Chalmers anticipated this extension of the extended mind thesis, noting that "In an unusually interdependent couple, it is entirely possible that one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto" (Clark and Chalmers 1998, p. 17). And there's no reason to restrict the social extension of cognition to just particular persons either. Some have argued (see, for example, List and Pettit 2011) that we can speak of the attitudes held by *groups* as well as individual people. If so, socially extended cognition can include group attitudes, especially those groups with which a person identifies.

At this point, the distinction between individual authenticity and group authenticity becomes very muddled indeed. Some of my beliefs, desires, or values will be held because they are stored in memory: they will be books in my mental library. Other attitudes I have will be virtual, but they will be *my* virtual attitudes: they will be entailed by various attitudes stored in my mind even if I am not presently aware of it. But then there will be other attitudes that are part of my socially extended mind: those attitudes I have because I identify with certain groups. These attitudes will be pointed to by the slips of paper found between the books in the library of the mind. And the attitudes that appear in my socially extended mind may *also* be held either actively or virtually. Conflict can occur between all of these.

What do we do when a conflict occurs? If I choose to abdicate an attitude in *my* mind in deference to a *group* attitude – a group forming part of my social identity – does that mean I am acting inauthentically? It's not obvious how to answer this question once we move away from the limited cases of obviously acting in bad faith or choosing in full knowledge after sufficient reflection. Suppose I believe that I don't have very good reasons for my individually held attitude and that much better reasons exist for the group attitude I adopt, even if I don't know what those reasons are. This doesn't seem to be an instance of not being true to myself; it seems, rather, to be a recognition of my cognitive limitations, an expression of epistemic humility in that I am willing to learn from those whom I consider to be better informed. In general, much knowledge acquisition features this social component. But now consider a more problematic case. Suppose that I do have good reasons for my attitude, but then I adopt the conflicting group attitude out of pressure to conform. One could argue that this is a violation of authenticity, of not being true to myself. But what if I did so out of *my desire* to conform? If I conform to social pressure because I am following a norm and – following Bicchieri's theory of social norms – I am the kind of person who prefers to conditionally conform to a norm on the assumption other people do, why should that mean I am not being true to myself? What justifies treating my conditional preferences as second-class preferences, counting less than my unconditional preferences?

Essentially, the trouble with WINOs and those who berate them as such is the following conflagration of issues. First, when a person's tribal identity matters to them, it leaves them vulnerable to being manipulated and exploited by those who purport to speak for the group. Those speakers do not need to be actual group leaders in any official capacity; often it suffices that their speech has enough reach to create the impression that they are instrumental in shaping the collective identity. Second, our willingness to conform to social norms – our preference to comply with certain forms and rules of behaviour *given that* enough other people do – means that the cost–benefit calculation we perform when a conflict arises between the individual and the group may, in cases of uncertainty, tend to tip in favour of the *group*. Third, given the importance of authenticity for our self-understanding, it's easy to see how the positive psychological feedback obtained by behaving in a way publicly recognised as authentic by a group – especially a group featuring prominently in our social identity – may quash any personal reservations one may have about giving up a personal commitment. Fourth, when we allow for the possibility of socially extended cognition, the very distinction between individual authenticity and group authenticity becomes blurred. In some cases, the two concepts of authenticity may be perceived by a person to effectively collapse into one. When this happens, we arrive at the paradoxical situation where the ability to be true to one's self is no longer solely under one's control: it is ceded to the forces determining one's tribal identity. The desire to live authentically can, in this case, become subverted by groupthink and conformity.

Other grounds for concern exist. Suppose that someone is called out for being a WINO, an inauthentic group member. Rhetorically, this is a violation of what, in the legal context, would be called the presumption of innocence and a shifting of the burden of proof. The charge of inauthenticity is made public, effectively requiring a response by the person alleged to be a WINO. This places the WINO at a disadvantage for three reasons. First, authenticity is generally viewed as an all-or-nothing state; like pregnancy, a person cannot be just a little inauthentic. Refuting a charge of inauthenticity requires establishing that the WINO has never made a decision or acted in a way that goes against those characteristics required of an authentic group member. Proving a negative claim is extremely difficult. Second, the charge of inauthenticity frames the discussion in a manner unhelpful to the WINO. The status quo has shifted, so that the conversation will proceed with listeners judging whether what the WINO says suffices to persuade them to move *away* from the charge of inauthenticity. This matters because psychological experiments have shown that status quo bias wields a powerful influence on human decision-making. Third, for the vast majority of groups and organisations, what it means to be an "authentic" member of the group is not universally agreed upon. Authenticity is thus a Rorschach concept upon which every group member can project his or her understanding. The WINO then faces the near-impossible task of persuading each member of the group that, according to his or her own understanding of the group identity, the WINO actually meets those criteria.

Those are some reasons why it's bad to be called a WINO, from the WINO's point of view. Let's now consider some sociological reasons why the very concept of a WINO, of appeals to group authenticity, can be harmful to society.

When someone is said to be an *authentic* $W$, what does that mean? Authenticity has many connotations, but some of the important ones are as follows. An authentic $W$ is someone whose credentials as a $W$ are undisputed. They are seen as a loyal and reliable member. They are judged as having been faithful to whatever principles and values are constitutive to the identity of the group. Authenticity, then, suggests a principled consistency.

There are many occasions when principled consistency is to be cherished. However, when group authenticity becomes fetishised to the point where inauthenticity becomes a term of critique, we should recall Emerson's observation: "A foolish consistency is the hobgoblin of little minds" (Emerson 2005 [1841], p. 40). Labelling someone a WINO is often an attempt to control the group, an attempt to reduce the diversity of viewpoints, and an exercise of power by those who claim authority to determine what makes someone an authentic group member.[4] Requiring group members to adhere to a single worldview can be counterproductive to the advancement of the interests of the group. Groups exist in a dynamic world, and the fact that established principles or ways of being have generally functioned in the past provides no guarantee they will continue to do so in the future. Allowing for natural variation within the group provides a degree of resilience against the unexpected. This is as true for cultural groups as it is for biological groups.

An obsession with group authenticity is often closely related to concerns about ideological purity. Yet ideological purity creates potentially disastrous conditions when two groups with incompatible worldviews interact. Real-world problems are complex and messy and often cannot be solved without compromise. If individual members find themselves self-censoring, or constraining their behaviour, out of concern of being branded inauthentic, that can reduce the possibility of successful compromise. Furthermore, inter-group relations can become increasingly volatile when framed in terms of authenticity. Why? Since authenticity is an all-or-nothing concept, the combination of inter-group relations and authenticity encourages conceiving the interaction in terms of a zero-sum game. Any deviation from authenticity, on either side, is seen as an unacceptable loss. Yet such deviations may be required to achieve joint concessions, yielding an overall net gain for each group. Such deviations will be difficult, if not impossible, to entertain if they involve attitudes understood as *necessary* or *essential* to be a group member. Yet there are very few social groups where what it means to be an *authentic* member is not socially constructed and hence, in principle, subject to revision.

A further trouble with charges of inauthenticity is that it merely attacks *what* the WINO says or does, rather than the *reasons* underlying the WINO's speech or behaviour. Sometimes there might be very good reasons for a person speaking or acting contrary to what one would normally expect from an authentic group member. Nixon's trip to China in 1972, despite being deeply out of character for the strongly anti-Communist Republican party of the time, was absolutely crucial to a thawing of relations between the US and China. Despite it being open to charges of inauthenticity from those who believed a Republican president should be a strong anti-Communist hawk, it was the right thing to do. Similarly, Dwight Eisenhower – the former Supreme Allied Commander in Europe during World War II, who led the D-Day invasion of Europe in Operation Overlord – ended his two terms as US President with a warning in his farewell speech about the excessive influence accruing to the military–industrial complex.

Finally, the elephant in the room about WINOs and those who complain about them is the basic problem of who decides what counts as authentic. The collective identity of groups emerges from the beliefs, desires, values, and actions of the people who belong to them. When a person, or a set of persons, attempts to brand someone as inauthentic in the absence of clear supporting evidence or reasons, that effort should be seen for what it is: an attempt to marginalise, exclude, or delegitimate the person. But disagreements over authenticity need not always mean that. Although some might find the existence of division within the tribe over what it means to be a $W$ frustrating, as long as the division exists because people are acting sincerely and in good faith, this disagreement should be seen as the cost of living in a free and open society.

Concerns about authenticity become especially vexed when we take into consideration the fact that a single person's social identity involves membership in multiple groups at the same time. How does a person decide what to

do when they belong to several groups and the constitutive rules for membership in those groups point to different courses of action? One of the difficulties with speaking of a person being an authentic $W$ is that almost no one is *just* a $W$: a person is both a $W$ *and* an $A$ (and a $B$, and a $C$, …) This brings us to the concept of intersectionality, which we will consider in the next chapter.

# 26. Intersectionality

In thinking through these issues about modern tribes, one assumption made throughout was that a person's tribal identity was relatively straightforward. Sometimes this is true. There are people who clearly identify as conservative or liberal, Muslim or Christian, who tick all the membership criteria and fit squarely within the tribe. But, in other cases, a person's social identity is less straightforward. This can happen for a variety of reasons. Sometimes it happens because a person feels that they are at the margin of a group and just don't fit in. Sometimes it happens because no social group which the person can see themselves as belonging to exists. Of these cases, the most interesting cases occur not when people simply haven't yet organised but because the relevant *concepts* have not yet been invented or discovered. The nature of transgender individuals, before "being transgender" entered into public awareness, may provide one illustration of this. And sometimes it happens because a person not only belongs to multiple tribes but because their existence at the intersection of multiple tribes yields an experience that is more than the sum of its parts.

Up to now, when talking about tribes the working assumption has been that people's tribal identity is *additive*. Although the obligations you might have as both an $X$ and a $Y$ could conflict – in the sense of competing demands on your time, or in the sense of Sartre's student who was torn between joining the French resistance or staying home to take care of his mother – the actual *experience* of being both an $X$ and a $Y$ was assumed to be just the result of combining the experience of being an $X$ and a $Y$ in isolation.

Sometimes, though, social identities are not merely additive. In these cases, what one experiences as a member of *both* $X$ *and* $Y$ is not the same as merely taking the experiences of being an $X$ and adding to it the experiences of being a $Y$. In these cases, the combination of the two categories creates unique experiences that are not shared by people who are $X$-but-not-$Y$ or $Y$-but-not-$X$. This is the phenomenon of *intersectionality*.

The emergence of the concept of intersectionality traces back to the experiences and concerns of Black feminists in the 19th century. First-wave feminism, which ran from the 19th century to the early 20th century, was primarily focused on achieving certain legal outcomes such as women's suffrage, the right of women to own property, and the right of women to have access to their children. Although these aims were ones that concerned *all* women, the fact that the social and political movements associated with first-wave feminism were largely composed of white, middle-class women meant that the concerns

of minority women were frequently sidelined. And while there is no *conceptual* reason why these two social groups necessarily need to be misaligned in this way, the specific historicity of the abolitionist movement and the women's rights movement in the US led to this tension.

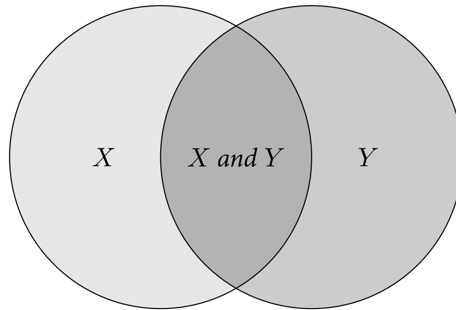
Despite the phenomenon of intersectionality existing for many years, the term "intersectionality" only entered contemporary social theory in 1989, when the legal scholar Kimberlé Crenshaw published an influential paper examining the ways in which the combination of gender and race created problems for the application of anti-discrimination policies in the law. Since then, the concept of intersectionality has been generalised beyond its original application in theories of feminism and race. This is why people now speak of intersectionality concerning the complexities involved by those who are $X$ and $Y$, for a wide range of values of $X$ and values of $Y$.

Intersectionality has become an important conceptual resource for understanding the dynamics of power, discrimination, and marginalisation. Calling attention to the non-additive nature of social identities and the complexities that they generate for the law, public policy, and social theory is extremely valuable. At the same time, a number of criticisms have been raised about intersectionality theory, such as a tendency to theoretically privilege identity at the expense of other sociological factors, like economic class. Some have also argued that supporters of intersectionality theory have, on occasion, allowed the perfect to become the enemy of the good. The moment one acknowledges the non-additive nature of being an $X$ and $Y$, it becomes easy to find apparent examples of exclusion where none was intended. Any gathering of $X$-types will, of necessity, fail to include a number of $X$ *and* $Y$ for at least some values of $Y$. Is this failure of inclusion an act of deliberate discrimination on behalf of the few $X$-types who gathered, i.e., is it an attempt to marginalise certain perspectives? Or is it simply because every person speaks from some perspective and any organised meeting has finite capacity? Helen Lewis noted:

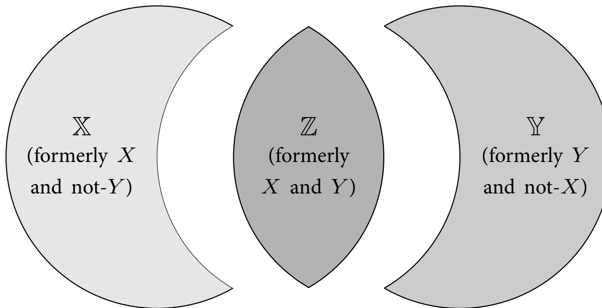
> When Caroline Crampton and I got together our bloggers last year for a *New Statesman* debate about feminism, the response was […] well, there were two responses. There was criticism that was *constructive* […] And there was criticism that was *destructive*, aimed at wounding us for not representing every possible permutation of womanhood. (I laughed when one particularly enthusiastic deconstructor, when asked: 'Well, how can you possibly make a six-person panel totally representative of half of humanity?', came back with, 'Oh, that's why I don't believe in panel discussions.') (Lewis 2014)

Given the social psychology of modern tribes, there's another point of concern worth noting. Suppose we have two tribes, $X$ and $Y$, such that membership in these tribes is not mutually exclusive. As we have seen, the composite

**Figure 26.1: The fracturing of tribes along the lines of intersectional identities**




(a) Before the redefinition of tribal identities: two tribes with some overlap




(b) After the redefinition of tribal identities: three tribes, with no overlap

Source: author.

identity of $X$ *and* $Y$ can create unique possibilities and experiences. Let us assume this is the case. It then becomes possible for those persons to convert their collective intersectional identity into a new tribe $\mathbb{Z}$, where members of $\mathbb{Z}$ are those who are both $X$ *and* $Y$. Once this happens, the new social understanding of what remains of tribe $X$, call it $\mathbb{X}$, are those who are $X$ *and not*-$Y$. And something similar holds for the new understanding of what is left of the original tribe $Y$. Figure 26.1 illustrates this fracturing of identities.

In reality, the fracturing of identities will rarely be neat and orderly. Some who are both $X$ *and* $Y$ may resist treating their intersectional social identity as a new tribe in itself. Others may feel liberated by such a fracturing, believing that only by declaring $X$ *and* $Y$ to be an independent tribe of its own can the $X$ *and* $Y$ avoid subordination. In this renegotiation of tribal identities, all the issues we've looked at previously reappear in new ways. Is someone who is $X$ *and* $Y$ but doesn't want to align themselves with a separate tribe – perhaps they are happy just seeing themselves as primarily an $X$ or as a $Y$ – not authentically

embracing their intersectional identity? What happens when an $X$ *and* $Y$ and an $X$ *and not-*$Y$ disagree over what it means to be $X$? For example, think about the debates between different schools of feminist thought regarding the nature of transgender women or sex workers. And how does our tendency towards in-group bias and negative stereotyping of out-groups influence us once an intersectional identity becomes understood as constitutive of a group *itself*, rather than as existing in the overlap of two groups?

This last point warrants consideration because when intersectional identities become constitutive of tribes, the intersectional component has the power to cleave. As previously noted, "to cleave" means both to divide and to join. If we consider the tribes in Figure 26.1, the intersectional identity of $\mathbb{Z}$ joins it with both $\mathbb{X}$ and $\mathbb{Y}$ through the overlap of common parts. But the intersectional identity of $\mathbb{Z}$ may, on occasion, divide it from $\mathbb{X}$, for members of $\mathbb{X}$ are also not-$Y$s whereas members of $\mathbb{Z}$ are $Y$s. Whether or not the identity of $\mathbb{Z}$s is seen as joining or dividing them from members of $\mathbb{X}$ thus depends on which aspects of $\mathbb{Z}$'s identity are salient. When the $Y$-aspect of a $\mathbb{Z}$'s identity matters, it is possible that members of $\mathbb{X}$ will be seen as the *out-group*. Similarly, when the $X$-aspect of a $\mathbb{Z}$'s identity is salient, members of $\mathbb{Y}$ may be seen as the out-group, in turn.

This ability for members of $\mathbb{Z}$ to conceive of those who are $\mathbb{X}$ or $\mathbb{Y}$ as the out-group, depending on which part of the $\mathbb{Z}$ identity is made salient, has two effects. First, it serves to increase the number of ways social conflict can occur. Second, the fact that members of $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ have their own tribal identity could reduce everyone's ability to engage in collective action, even when interests are aligned.

The increase in the number of ways possible social conflict can occur is easily seen. Before the intersectional identity became constitutive of a tribe, there was only one source of possible social conflict: between the tribes $X$ and $Y$ (remember, this is before the fracturing) regarding those who were outside the intersection. After the fracturing, there are three possible sources of social conflict: between $\mathbb{X}$ and $\mathbb{Z}$, between $\mathbb{Y}$ and $\mathbb{Z}$, and between $\mathbb{X}$ and $\mathbb{Y}$. Although these possible conflicts will not always be realised, the change introduces a new source of social instability. As we saw with the Red Devils and Bull Dogs, it takes relatively little friction between groups to plant the seeds of animosity, even when those groups share a common starting point.

The ease with which group animosity can emerge also connects to the reduced efficacy of collective action. Suppose, for example, that there is some fourth group $\mathbb{W}$, whose interests are entirely opposed to those of $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$. (Assume, for the sake of argument, that there is no moral reason as to why $\mathbb{W}$ has a claim to advance its interest; everything we are talking about concerns the self-interest of all four groups.) Initially, this might seem to take the form of a superordinate goal problem, of the kind discussed in Chapter 24. Yet there is an important complication. A problem of *strategic choice* is one where people have to decide the best way to proceed in an environment containing other agents who are trying to decide the best way to proceed, where the notion of

the "best way to proceed" is a function of everyone's choices. The introduction of the fourth tribe $\mathbb{W}$ here means that $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are confronted by a problem of strategic choice against $\mathbb{W}$. In particular, there is a zero-sum game between $\mathbb{W}$ and $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$. This transforms the nature of the superordinate goal the $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ tribes have to address.

Although all three tribes have an interest in working to mitigate the impact of $\mathbb{W}$ upon them, the potentially competing tribal identities of $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ affect how the collective action problem may be framed. Let us say that the three tribes have *aligned* interests but not *common* interests. Why not common interests? Recall the discussion of common meanings from the introduction: a common interest requires a sense of community that binds the individuals together. A common interest requires people to be able to say, "This is a concern of *ours*, and it is in *our* interest to address it." The ambiguity of "our" means it can be read either widely – referring to all three tribes with the same interest – or narrowly – referring just to the tribe to which the speaker belongs. It is the narrow reading that proves problematic.

Consider the collective action problem from the point of view of members of $\mathbb{Z}$ deliberating over what to do. In most real-world situations, the distribution of benefits resulting from collective action problems will be unequal across both groups and members. Suppose that there are several proposals on the table for how to resolve the conflict between $\mathbb{W}$ and $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$. Suppose that some outcome benefits members of $\mathbb{Z}$ but, at the same time, benefits members of $\mathbb{Y}$ more. If the $X$-aspect of $\mathbb{Z}$'s identity, for some reason, becomes particularly salient, members of $\mathbb{Z}$ might wonder why it is that the $\mathbb{Y}$'s have received a better deal. Is one reason the $\mathbb{Y}$s are so keen to participate because they come out on top? The $\mathbb{Z}$s may then discuss whether it would be better to adopt some other course of action, one that allows them to do a little bit better and, perhaps, favours the $\mathbb{X}$s rather than the $\mathbb{Y}$s. If the $\mathbb{Z}$s proposed this alternative course of action, it would now be the $\mathbb{Y}$s' turn to be irritated. Perhaps it would be better for the $\mathbb{Y}$s to work more closely with the $\mathbb{X}$s, leaving the $\mathbb{Z}$s out in the cold. The point is this: by assumption, it's in the interest of $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ to cooperate against their common enemy, but the possibility of intertribal disputes between these three gives $\mathbb{W}$ the ability to cultivate and exploit these divisive tendencies, thereby ensuring that the opposition remains less effective.

Thus we find a counter-intuitive result: intersectionality theory, originally developed as a conceptual tool to combat discrimination and address social injustice, may – when intersectional identities become concrete tribal identities through reification – trigger social psychological processes that impede the pursuit of social justice. This concern acquires even greater force when we realise that the possibilities for intertribal conflict are more complex than Joshua's question acknowledged. Joshua asked, "Art thou for us, or for our adversaries?" The key assumption was that only two options exist: you are either for us or for our enemies. In short, my enemy's enemy is my friend.

This way of thinking about group relations underlies a *realpolitik* approach to the exercise of power. If, however, one is concerned about who may *legitimately* exercise power, this binary logic no longer holds. In this case, my enemy's enemy may still be my enemy. The mere fact one group is opposed to a second group, where the second group cannot legitimately exercise power, does not itself confer legitimacy upon the first.

Putting all this together, we can see how tribal identities, when combined with identity politics, can constitute a threat to the communitarian conception of the Open Society. Individual identities are multidimensional, and these multiple dimensions provide fault lines along which society may fracture into competing tribes. These competing tribes, when concerned with the legitimate exercise of power, can find it hard to engage in effective collective action because doing so requires cooperation with those who are seen as morally tainted. This situation can be exploited by political actors willing to engage in *realpolitik* and use social division to consolidate their grip on power.

This, then, is the real challenge posed by Joshua's question: how do we understand the "us" in the question, "Art thou for us?" Social identities centred on abstract, inclusive characteristics are less compelling for motivating action than social identities centred around concrete, specific characteristics. Yet the more concrete and specific the social identity is, the greater the possibility of fracture due to the psychology of social identity.

# 27. Epistemic closure and extreme groups

In 2004, the *New York Times* journalist Ron Suskind interviewed an unnamed official in the Bush Administration who criticised the "reality-based community". According to the source, the reality-based community comprised those people who "believe that solutions emerge from your judicious study of discernible reality". The source went on to say: "That's not the way the world really works anymore. We're an empire now, and when we act, we create our own reality."

A charitable interpretation of what the official said would not attribute to him an outright denial of the existence of reality or objective facts. Rather, the actions of a major player like the US didn't simply occur *in* a geopolitical landscape, they *shaped* the landscape in the first place. Looking back, it's now clear the official overestimated the ability of the US to create its own reality in Afghanistan and Iraq, but that hasn't stopped other authoritarian leaders from trying to do the same in other parts of the world.

Those in the reality-based community are right to insist on "the judicious study of discernible reality" for a number of decisions. There's no mental state a person can put themselves in which will render polonium safe for human consumption or novichok non-toxic. The physical world places non-negotiable constraints on what people may reasonably believe about it. We neglect those constraints at our peril.

When it comes to the social world, things are considerably different. Social reality is underdetermined by the physical reality existing outside of peoples' heads. Much of social reality exists in the way that it does simply because people have collectively decided on certain things. Beliefs, norms, and conventions are passed down from parents to children, and the fact that human generations have long periods of overlap with extended periods of education ensures a reasonable degree of continuity in the understanding of social reality.

But the fact that social reality is not *determined* by the physical facts outside of a person's head means that, from time to time, a group's understanding of the social world can become, from the point of view of those on the outside, radically unhinged. The considerable independence of the social world from the physical world means that beliefs that strike us as bizarre or abhorrent can not only endure but can be surprisingly resistant to change. Think of racists who believe in their own ethnic superiority, of religious cults, of violent terrorist movements like Boko Haram, al-Qaeda, and Islamic State, and of the "involuntary celibates", also known as incels, who have committed several mass shootings in recent years. This brings us to the darker side of modern tribes:

those groups that are harmful or dysfunctional, whose view of the world strikes us as utterly mistaken, but which, for some reason, aren't self-correcting.

This is the phenomenon of *epistemic closure*: the formation of a system of beliefs and values that appears, to those of us on the outside, to be disconnected from the normal relations holding between information, evidence, and belief formation.[1] An epistemically closed system is one that has certain beliefs and values such that, regardless of the evidence, the attitude of a person embedded in that system towards those beliefs and values won't change. Systems can exhibit partial epistemic closure as well. When a system of beliefs is partially closed, information and evidence are much less effective at shifting a person's attitude towards those beliefs than one would expect if people were adjusting their beliefs as purely rational agents.

The possibility of epistemic closure is, unfortunately, a fact of life. Any empirical theory, or any set of general beliefs about the world, is underdetermined by evidence. There's no number of observations that one can make which will suffice to establish the truth of an empirical claim of any generality unequivocally.[2] The matter is even worse when we consider non-empirical beliefs, such as ones involving metaphysics and religion. If a person decides to maintain an empirical belief that runs counter to the available evidence, it's always possible for the person to do so. Of course, if they wish to keep a consistent and coherent set of beliefs, resisting the evidence comes at a cost: the person will have to make increasingly complex and odd adjustments to *other* beliefs they hold in order to maintain the belief they wish to hold fixed.[3] A classic example is how advocates of the geocentric model of the solar system had to add epicycles to planetary orbits to make it fit with observations. But there is a positive lesson to be drawn from the underdetermination of theory by evidence: if we can't ever truly be certain about our empirical beliefs, we shouldn't be dogmatic about any of them either; we should always treat any belief about the world as revisable in principle. This recognition of our fallibility should, if anything, lead us towards a position of epistemic humility. So what is it about some groups that causes them to become epistemically closed?

Epistemic closure can occur for a variety of reasons. One mistake people often make is to treat beliefs – cognitive attitudes towards propositions that have a truth value – as if the *only* thing that matters about a belief is its truth value. Take the belief that anthropogenic climate change is occurring. This belief has a truth value independent of whether we actually know its truth value. One common model of how people form beliefs is that, in an ideal setting, they take evidence into account when forming a judgement. In such an ideal setting, people would consult the findings of climate science or atmospheric science more generally. However, moving from idealisations to the real world, we encounter the problem that most of us cannot engage with the relevant scientific literature because we don't have the expertise or the time required to understand the details. Given this, we do the next best thing: we rely on another person's expert opinion to form a judgement by proxy, using *their* determination to fix *our* belief. Or, since most of us don't have a single scientific

expert on whom we rely, it's more likely that we rely on a trusted *epistemic community* to fix our beliefs. Traditionally, in the case of scientific beliefs – where we are motivated solely by *truth* – the trusted epistemic community will be the scientific community, since their expert knowledge and track record give them the credibility to act as arbiters regarding such matters. In this idealised model, what we keep returning to, time and again, is the importance of truth for the fixation of belief.

What the above story leaves out is the fact that, for many people, what matters most about a belief is not its truth value but the relationship between the person and a tribe associated with the belief. The belief becomes a signifier of tribal membership, and endorsing the belief often becomes an informal requirement of tribal membership. The idealised model conceives of the matter this way: beliefs should be true and, therefore, when I cannot fix a belief with sufficient reason to be confident that it is true (or likely to be true), I should rely on the appropriate epistemic community whose track record gives me assurance they will not lead me astray. The contrary model conceives of the matter this way: my beliefs situate me within a particular tribe, and that tribe contributes to my social identity. My social identity – as a source of esteem, friendship, camaraderie, and how I understand my social existence as well as the determiner of meaning for my life – is sufficiently significant that preserving my social identity is a matter of fundamental importance. Since I am not in a position to form a judgement on my own, I need to rely on other people to guide my belief formation. Since my tribe (for all the reasons just mentioned) has been a key set of people who show me respect, understanding, friendship, and so on, I rely on them to guide me. Consequently, I defer to the epistemic community *determined by my tribal identity*[4] in the fixation of my beliefs because doing so is the best way to ensure continued membership in the tribe.

This inversion of grounds for belief might seem weird. To begin, it would seem to decouple beliefs from reality in ways that the "reality-based community" would legitimately see as harmful. If there is no mental state a person can adopt which, for example, would render polonium safe to consume, why wouldn't blind deference to beliefs – especially ones which we on the outside might describe as "batshit crazy" – determined by a tribal identity be, if you like, eventually eliminated by natural selection? How can obviously false, individually harmful, and socially damaging beliefs persist?

To be sure, if an extreme group's beliefs become sufficiently misaligned with reality, the group will eventually suffer. When the beliefs of the Manson Family led them to murder Sharon Tate and four other people, reality caught up with them through the efforts of the police and they were imprisoned. The Heaven's Gate religious cult, whose members believed that they could escape Earth by committing mass suicide in order to join a spaceship that was following Comet Hale–Bopp, has no followers today. The followers of the reverend Jim Jones, who committed mass suicide at Jonestown, Guyana, also paid the ultimate price for their beliefs. Many other examples exist.

Yet many other deeply flawed beliefs persist amongst groups but are not driven out. Here we must recognise that, in the developed world, we have engineered society in such a way that we rarely need to rely on the majority of our beliefs in order to navigate the world safely. If you have what Graeber (2018) calls a "bullshit job", you don't need many beliefs to do your job and get paid. If your most deeply held beliefs go against what you are required to do in your job, you can go through the motions bracketing what you really think and tell yourself, "*This is what I need to do in order to get paid.*" A person doesn't need to have accurate scientific beliefs, accurate political beliefs, or accurate economic or social beliefs to go to the supermarket and buy food. A person doesn't have to believe in evolutionary theory to go to the doctor, be prescribed antibiotics, and take them to get better. A person doesn't have to believe in general relativity to get directions via GPS, even though the technology would malfunction if the designers didn't take into account the gravitational time shift implied by general relativity. Sure, to someone fundamentally committed to truth it appears deeply *hypocritical* to behave in such a disingenuous fashion but so what? The person who acts this way will still feed themselves, be cured of their infection, and be able to get to where they want to go, regardless of what they believe. The only beliefs a person really needs to survive in a modern society are beliefs like how to cross the street without being run over, how to pay your bills on time, how to cook over a gas stove without blowing the house up, and so on. These beliefs are highly specific pragmatic local knowledge and are compatible with a wide variety of highly unorthodox theoretical beliefs. Society's technological prowess has, for better or worse, radically decoupled people's ability to survive from the theoretical coherence, truthfulness, and accuracy of their beliefs.[5]

This decoupling of a person's ability to survive from the truthfulness and accuracy of their beliefs means that a person's system of beliefs can acquire a different functional role. What may matter most about a belief is not its content or *actual* truth value but what holding that belief signals about a person and their tribe. The denial of anthropogenic climate change provides a nice illustration of this phenomenon in action. Climate change requires a coordinated, global response. The actions of any single person are entirely irrelevant to the global outcome. Due to the long delay in the climate's response to environmental legislation, it is unlikely any person will see a connection between their actions and change to the environment over the short to medium run. Given this, beliefs about whether anthropogenic climate change exists can be co-opted to serve a signalling function about which tribe a person belongs to because, over the short run, they are decoupled from noticeable material consequences. This seems to be what has happened with the Republican party in the US. Under Nixon, as noted previously, the Republican party passed a number of pro-climate pieces of legislation. Today, many Republicans are sceptical about climate change. Why? The importance of their social identity *as Republicans* leads them to defer to the epistemic community identified as credible *by their tribe*. The real question is why has the Republican tribe identified, as

credible, an epistemic community that denies climate change? Because many economic and business interests represented among top Republican donors benefit from continuing with business as usual, rather than making the effective changes required to combat climate change. The epistemic community recognised as credible by the Republican tribe has been ideologically captured and subordinated by these economic and business interests. Furthermore, the denial of climate change can be spun in two ways. First, as a rejection of "liberal science" with its purported political bias. Second, as an attempt to bring back traditional extractionary industries (such as coal mining) or new ones (such as fracking), for which there is an interest in communities that have historically voted conservative. In both cases, climate change denial reaffirms the tribal identity of those who adopt the belief. Truth takes a back seat to these other social functions.

This last example illustrates how the various social functions played by beliefs can subordinate the truth-functional role of beliefs. Once we recognise that there are alternative functional roles played by beliefs that can trump a concern for truth, we see that confronting the phenomenon of extreme groups, and the epistemic closure that accompanies them, requires engaging with the underlying *functions* that are served by the system of beliefs, rather than engaging with their theoretical content. And this shift towards the functionality of systems of beliefs means we have to acknowledge that, sometimes, the real social function served by a system of beliefs is not necessarily known by many – perhaps *any* – of the people who have those beliefs.

This is the well-known distinction between *latent* and *manifest* functions, deriving from classic work in anthropology by Malinowski (1941) and Radcliffe-Brown (1952), among others. A manifest function is a function a social practice has been consciously designed to have, such as how randomised police patrols keep criminals from being able to predict a safe time to commit burglaries. A latent function is a function a social practice has but for which it was not consciously designed. A classic anthropological example is how the practice of extended lactation in hunter-gatherer tribes (i.e., breastfeeding infants for longer than 12 months) had the latent function of controlling the population because breastfeeding reduces fertility.

When it comes to the social functions of belief systems, the distinction between latent and manifest functions helps us understand the phenomenon of epistemic closure. Epistemic closure seems irrational when we only consider beliefs as vehicles of truth. When we see that belief systems can have, as their *latent* function, the satisfaction of other psychological and sociological needs, the fact that those beliefs are unresponsive to evidence no longer seems unusual. If a belief system helps me understand my place in society, explain why I am unhappy (or happy) or unsuccessful (or successful), and legitimates how I feel, there's little incentive to change those beliefs because doing so leaves me with a gaping lacuna of unaddressed psychological and sociological needs.

Take, for example, why so many rich people believe in efficient markets. It's not because the models of market efficiency are deeply compelling: the conditions under which we can show markets converge to an equilibrium such that supply equals demand for every commodity in the economy are pretty demanding. Believing in the efficiency of markets isn't justified because of the goodness of fit between the theory and the reality of the modern economy. Rather, I suspect rich people often believe in market efficiency because the general equilibrium theory, and the existence of perfectly competitive markets, provides a putative meritocratic justification for their wealth. Few people want to admit that they are rich because they got lucky or abused an underregulated monopolistic position, or exploited a vulnerable workforce with little bargaining power. The theoretical content of the belief *markets are efficient* is what really provides the psychological function *I deserve to be where I am*.

Keeping in mind the idea that belief systems can serve latent social functions, now consider one extreme group that has received a considerable amount of media coverage recently: incels. In speaking about "incels", I am using the more recent understanding of the term, which refers to a movement centred around young men who frequent a number of online forums featuring misogynistic ideas, including violently punishing women for not having sex with them.[6] The incel movement has been identified by the US Southern Poverty Law Centre as a worrying example of male radicalisation online. Incels came to international attention in 2018 after Alek Minassian killed 10 people in Toronto in a van attack, which he claimed in a Facebook post prior to the event as his contribution to the forthcoming "incel rebellion". Minassian's action followed in the footsteps of the mass shooting by Elliot Rodger in 2014, in which he killed six people (seven, if you count Rodger) in California. In a Facebook post, Minassian glowingly referred to Rodger as the "Supreme Gentleman".

The media reaction to the Minassian rampage, by and large, condemned the Toronto attack as a murderous assault by a member of a radical fringe group. Some engaged thoughtfully with the wider cultural backdrop in which the phenomenon of the incels occurred, arguing that, although no one has a right to sex, we as a society need to reflect on the ideals of beauty that are disseminated throughout the culture, which lead to the rejection or marginalisation of those men and women who fall short of the ideal (Srinivasan 2018). But other commentators took the rhetoric of the incels at face value, wondering whether people really do have a right to sex and whether we ought to consider methods of redistributing sex. The *New York Times* columnist Ross Douthat attracted a fair amount of online opprobrium for broaching the question, "If we are concerned about the just distribution of property and money, why do we assume that the desire for some sort of sexual redistribution is inherently ridiculous?" (Douthat 2018). Here his comments were similar to those of the economist Robin Hanson, from George Mason University, who posted the following on his blog a few days earlier:

> One might plausibly argue that those with much less access to sex suffer to a similar degree as those with low income, and might similarly hope to gain from organizing around this identity, to lobby for redistribution along this axis and to at least implicitly threaten violence if their demands are not met. (Hanson 2018)

These comments strike me as misguided, managing to intersect with questions worth asking, but in a way that is, at best, only indirectly related to understanding the phenomenon of misogynistic incel culture.

The online culture of incels, and the system of beliefs and values associated with it, provides a therapeutic function for young men who have not succeeded in realising, in whole or in part, a vision of their life as they wished. The conspiratorial narrative of "Chads and Stacys" who thwart their life plans provides an external locus of control, absolving incels of responsibility for their situation. And although some of the ideas and unsatisfied preferences that push people to self-identify as incels derive from a culture of toxic masculinity, of unrealistically promiscuous pornographic sexuality, not all do. The idea, for example, that there is someone for everyone is deeply embedded in popular culture, along with norms about how heterosexual relationships will develop over time. If you are a man unable to find work that pays well enough to support a partner or a family and, for whatever reason, also find yourself unable to form relationships with women, the contrast between your life experiences and those of "successful men" – as stereotypically represented within the wider culture – will be unsettling.

Given the violent, misogynistic nature of online incel commentary, it is natural to want to engage with the surface meaning of what is said, to criticise and condemn. Yet this is to mistake symptom and cause. There is a need to distinguish the normative question of whether men *should* have certain preferences and expectations from the descriptive questions of *what* are the expressed preferences and expectations, *why* they have those preferences in the first place, and how men behave when those preferences and expectations are not satisfied. It is only by identifying, and targeting, the latent function served by the incel system of beliefs that we will be able to, as a society, make progress towards eliminating the hate and misogyny by addressing the real underlying cause.[7]

The idea that systems of beliefs can provide a therapeutic function goes beyond that of just incels. In *Healing from Hate*, Michael Kimmel argues that one common factor contributing to the radicalisation of men, whether it is Neo-Nazi groups or Islamic extremism, is the failure to realise a certain ideal of masculinity. Feeling that their life has no purpose and that they have no way to live up to the gender role they feel is expected of them, men join these organisations to try to realise the ideal in another form. Kimmel writes:

> There's a reason most of the extreme right is male: it's that masculinity is centered on something to prove, a quest. Perhaps, you

> might think, women have to prove something also; perhaps they 'prove' their femininity by attracting a man, getting married, and becoming a mother. Maybe, but it doesn't have the same propulsive force as proving one's manhood. Men must prove their masculinity to other men, in the homosocial arena of other men. Their masculinity must be credited, validated, affirmed by their peers. Historically, of course, they've proved their masculinity in the traditional time-honored way of their ancestors: in the workplace, as breadwinners. They've provided for their families, protected their homes, and defend their homelands. Take those roles away, and they have to find a new arena in which to prove their manhood. (Kimmel 2018, pp. 45–6)

In stressing the importance of attending to the latent functions served by extreme belief systems, I do not mean to suggest that the propositional content of those belief systems never matters. My aim in this chapter was rather to suggest that a more nuanced approach to how we understand extreme groups is occasionally warranted. It's very convenient for politicians to write off extremists as evil people who believe evil things. The problem with this line of thought is that the process of radicalisation is more subtle than that. How and why do people get recruited into extreme groups when they do? And when people become radicalised, what is the appropriate response to de-radicalise them? These questions can only be answered by looking beyond the propositional content of their beliefs.

People desire to make sense of the world in which they find themselves. People want to be able to tell a story about why their lives unfolded the way that they have. When things go well, people want to be able to take credit for what they have achieved; when things go poorly, people want to be able to save face. Given the flexibility in how people may theorise about the social world while still being able to navigate it successfully, stories that fit with how people want to understand their lives can be more valuable to them than stories that accurately describe the way the world is. When those stories become deeply incorporated into a person's self-understanding, epistemic closure may occur. The moral of the story is this: often it isn't a question of *what* a person believes so much as *why*, both above and below the surface.

# 28. The collision of horizons

Perhaps the greatest challenge facing the communitarian conception of the Open Society is how a diverse population deals with each other when they have incompatible beliefs and incompatible conceptions of how life should be lived. The incompatible beliefs might concern appropriate attire for men and women (e.g., the debate over whether to ban the burqa or what is an acceptable Halloween costume), sexual preferences (e.g., the acceptability of homosexuality), the nature of human relationships (e.g., civil partnerships, gay marriage, or polyamory), religious beliefs, political beliefs, forms of economic organisation, and many others. In a diverse, multicultural society, members of different tribes holding incompatible beliefs will encounter each other in public spaces and often encounter ideas with which they fundamentally disagree. What ground rules can be established for negotiating the collision of horizons when two people meet and each thinks the other is profoundly wrong?

The first point to keep in mind is that, despite appearances to the contrary, there is more people have in common than not. We just don't often acknowledge that fact. Evolutionary forces have shaped us for hundreds of thousands of years to be sensitive to group differences and tribal threats posed by the "Other". In contrast, humans have been living in societies with populations sufficiently large to enable regular anonymous interactions between individuals only since the invention of agriculture slightly more than 12,000 years ago. That inbuilt suspicion towards those who we perceive differently is exacerbated by the news and other forms of media: it's a fairly steady diet of conflict and disorder. And that's before we take into account all the people trying to *create* division. The attention economy of social media often exposes users to inflammatory content for the sole purpose of provoking a reaction, requiring them to engage more with the platform. The fact it also stokes fear and sows division is a side effect.

But the commonalities we share are significant. People want to live safely in a society governed by the rule of law. They want laws to be fair and transparent and enforced equally across all persons, even if there is disagreement about what means for a law to be *fair*. People want trustworthy and reliable social institutions on which they can depend. People want to have a place to call home, which offers comfort, of which they can feel proud. People want to have a form of work that pays enough for the necessities, as well as a little extra. People want a form of work in which they can take pride and feel that they are doing something important and meaningful. When not working, people want to be free to choose how to spend time pursuing other projects and

hobbies that they consider to be valuable. In addition to the pursuit of work and personal activities, people also want to have meaningful relationships with other people and feel that they are valued and appreciated by others. People want a sense of purpose and a sense that they are participating in something meaningful over the course of their lives. How exactly these goals, and others, are pursued and realised can take on a variety of forms and create the possibility of conflict, but suitably redescribed, we can see that all these goals are part of the common, shared human experience. People may arrive at different, incompatible answers to the same fundamental questions.

The second point is that different conceptions of how to live mostly lead to conflict when one or more groups believe that *their* form of life is the only correct one and they have a duty to make others conform. This attitude is dangerous. Given how peacefully Protestants and Catholics coexist in the US, and most of the world (Northern Ireland is a noteworthy exception due to its fragility), it is hard to believe that millions of people died during the European wars of religion in the 16th and 17th centuries. The difference in religious worldview, so destabilising back then, has now been largely incorporated into the background warp and weft of society. Tolerance has brought peace.

Similarly, one of the great achievements of the EU is how difficult it became – at least until Russia invaded Ukraine – for people to conceive of war between European nations. This difficulty does not mean that conflict is impossible; we know from the ethnic cleansing in Serbia, and the invasion of Crimea and Ukraine by Russia, that such a threat exists and is all too real. But, as Steven Pinker has argued, the better angels of our nature have been on the march for some time. Familiarity and interdependence, the need to combat common threats such as climate change or terrorism, all combined with the realisation that there is much more that unites Europeans than what divides them has served to reduce the threat of conflict.

The third point is that cultural diversity, instead of being a threat, is needed for robust, stable societies. Some think that societies with great diversity are fragile or unstable because the variation makes it difficult to agree on policy. Although greater diversity can make it hard to reach agreement quickly, it is worth asking why reaching agreement quickly is necessarily a virtue. There are times when reaching decisions rapidly is essential (e.g., natural disasters), but that is not the only decision-making context which matters. A monolithic culture where everyone had the same beliefs and values would reach agreement quickly, but it would also be vulnerable to bad ideas which fit in alongside the rest of the universally held beliefs. Variation in opinion helps ensure that a society cannot be brought down by a single popular yet fundamentally misguided idea or theory. A diverse culture provides epistemic inoculation against bad memes. *Cultural* diversity in beliefs and values provides benefits analogous to those provided by genetic diversity in a *biological* population.

Putting these three observations together yields an answer to our question: what is to be done, in a diverse society, when incompatible worldviews find themselves in close proximity with one another? The answer involves ideas

associated, somewhat ironically, with both Mao and Popper: let a hundred flowers blossom. (It must be acknowledged that Mao's support for this attitude faded quickly.) In order for the Open Society to survive, it must protect itself from being undermined from within by having its freedoms and tolerance used against itself. Such undermining occurs when a radical group takes advantage of high levels of tolerance to increase its membership, then exploits its size to restrict the freedoms of others, denying others the ability to go their own way. Given this, the Open Society must impose the following minimal constraint on the freedom of individuals: people can adopt beliefs and values as they wish, behaving as they choose, so long as they abide by Mill's Harm principle. And that means that they act towards out-group members with tolerance, granting other people the freedom to associate with those they choose and to pursue a meaningful life constructed on their own terms.

A society where inter-group relations are governed by the Harm principle and the principle of being intolerant of intolerance would seem to generate a paradox: when one group's form of life involves behaviour or conduct that contradicts the beliefs and values of another, there seems no way to simultaneously satisfy both principles. The elimination of harm would seem to require that the first group curtail those practices that offend the sensitivities of the second group, yet requiring the first group to curtail their practices shows intolerance in violating their freedom to live a life constructed on their own terms. How are we to square the circle?

To begin, we need to recognise that the Harm principle does not apply universally to every kind of harm regardless of its origins. It is important to distinguish between illegal harms intentionally inflicted upon others and those harms experienced when a person puts themselves in a position where they could have anticipated they would be harmed by the legal behaviour of others. Harms of the first kind are regulated by the Harm principle whereas harms of the second kind may not be. For example, a person has legitimate grounds for complaint if they were walking down the street and saw a flasher; that's the reason we have laws prohibiting indecent exposure. However, that same person would not have grounds for a complaint if they knowingly and voluntarily went to a known, documented, and authorised naturist beach and became offended by the sight of people lying *au naturel* in the sun.

In a diverse, multicultural society, it will not be unusual for some groups to adopt a form of life that has the potential to offend others.[1] It will prove helpful to distinguish between one group *causing offence* to another and one group *taking offence* of another. To say that group $X$ *causes offence* to group $Y$ is to attribute a causal relationship between some, perhaps all, members of $X$ and some, perhaps all, members of $Y$. This can happen in a variety of ways: one member of $X$ might verbally insult a particular member of $Y$ or a number of $X$s might engage in a disparaging chant attacking the group identity of the $Y$s in a public forum, without aiming to insult any particular person. These are clear examples of how members of one group can cause offence to another. But not all instances of causing offence are intentional. Suppose that a member

of $X$ violates an important social norm of the $Y$s without knowingly doing so, such as when an ignorant tourist wears shoes inside an empty Buddhist temple, where there was no one to inform them of the social norm violation. In such a case, that member of $X$ could cause offence to members of $Y$ accidentally. Whether or not the offence caused was intentional or not may affect our judgement of its severity or the blameworthiness of the agent. Even an unintentional causing of offence could warrant sanctions, if the agent was culpably ignorant and they *should* have known, for example. But in all instances, there is a common causal structure at play: an intrusion into the sphere of existence *of* the offended person *by* the person or people who caused the offence. When an offence is caused, it happens at the intersection of two different spheres of existence, with an interaction between members of different groups.

In contrast, group $X$ may *take offence* at another group $Y$ without any causal interaction between their spheres of existence. Members of group $X$, simply by virtue of knowing that the $Y$s have a social practice of which the $X$s disapprove, might be outraged or upset or annoyed or offended. (Recall the much-quoted, humorous definition of Puritanism: "The haunting fear that someone, somewhere, may be happy", see Mencken 1982.) The important difference between *causing* offence and *taking* offence is that it is possible for group $X$ to take offence at group $Y$ when group $Y$ was simply trying to go about living its life according to a manner of its own choosing. And while the Open Society needs to protect one group from unduly imposing negative externalities on others, as discussed in Chapter 3, mere existence is not a negative externality.

The distinction between causing offence and taking offence offers a resolution to the apparent paradox of how an Open Society can allow for diverse groups, with incompatible worldviews, while at the same time being governed by the Harm principle and the principle of being intolerant of intolerance. In public spaces, where people's spheres of existence intersect due to their shared location, people need to abide by norms of civility and respect and minimise the chances of causing offence. At the same time, people need to be charitable in their judgement of others, so that if an offence is caused, they react proportionally by taking into account whether it was intentional or not. When we consider people's behaviour in private spaces, the principle of being intolerant of intolerance means that if one group *does* take offence at the practices of another, those offences are not regulated by the Harm principle. Groups are free to live life on their own terms, and the mere fact that another group disapproves and takes offence is not sufficient reason to curtail its freedom of self-determination.

A similar point holds if one group takes offence at another simply by virtue of encountering the second group in a shared public space (assuming, of course, that the second group conforms to basic norms of civility and respect and is not provoking the first group). For example, consider the discussions in a number of Western countries over whether to ban the burqa. Much of this debate derives from people taking offence at the practice of wearing the burqa

as a result of seeing people wear it in a public space, where the people wearing the burqa are, arguably, not trying to *cause* offence. It's just that someone who chooses to wear the burqa cannot go out in public without other people seeing them, and other people might take offence as a result of merely seeing a woman wearing a burqa. In this case, the harm in the group that takes offence at seeing someone in a burqa is not a harm that falls under the Harm principle. The way I think the Open Society ought to approach the issue is as follows: any group that *requires* women to wear the burqa against their will is oppressive – and thus engaged in a practice that should be curtailed in order to protect the freedom and autonomy of women – but if a woman chooses freely to wear the burqa, then that is an expression of her personal freedom and unproblematic.[2] Furthermore, as an outsider to the group, the choices made by group members over how they live their lives are really none of my business. We must respect people's freedom to live life as they choose. As long as people are choosing freely and properly informed, let a hundred flowers blossom and be intolerant of intolerance.

One complexity with this resolution of the paradox derives from the fact that the conceptual distinction between public and private spaces is imperfect. What happens when a public space is occupied by members of one group when a few members of another group arrive? We now have a case of intersecting spheres of existence, but to say that the first group might *cause* offence to members of the second group, simply by continuing to be as they were, runs the risk of creating the tyranny of the minority. In such cases, a good rule of thumb for members of the second group to abide by is the following: don't take offence when none was intended. At the same time, a good rule of thumb for members of the first group to abide by is this: when spheres of existence unexpectedly intersect, a little goodwill and empathetic understanding go a long way.

# 29. Concluding remarks

What, then, is the overall assessment of the communitarian conception of the Open Society? We have seen how polarisation, in-group bias, and conceptions of authenticity can create problems for diverse societies. In this part, I have tried to both describe the phenomena and trace some of the negative consequences they create. In the next, and final, chapter of the book, I will try to suggest what we can do to try to overcome some of these issues.

One difficulty, I think, with reconciling communitarianism with diverse societies is that we don't frame the challenge in the right way. Part of that framing challenge lies in the fact that one popular guide to action – the Golden Rule – doesn't necessarily work well in diverse societies. The Golden Rule derives from the book of Matthew, chapter 7, verse 12, which in the King James Bible reads: "Therefore whatever you desire for men to do to you, you shall also do to them." In more modern language: "Do unto others as you would have them do unto you."

The Golden Rule has two problems. The first is rarely noted: it is compatible with antisocial preferences. If you are perfectly happy with other people being a jerk to you, then you should act like a jerk to them. This feature of the Golden Rule rarely surfaces because, most of the time, it is assumed that people don't have antisocial preferences. The second problem, which is more serious, is that the Golden Rule assumes that the best way to infer something about how *other* people want to be treated is to reflect upon how *you* want to be treated. In a homogeneous society this might work well, but in a diverse, multicultural society it can go awry. And although there will be a lot of agreement on how people want to be treated, at least at the level of abstract descriptions – i.e., to be respected, to be treated fairly, to be thought of as a good person, and so on – those abstract descriptions can be realised in a number of ways.

To resolve these problems, I think the Golden Rule needs to be replaced with two rules better suited for diverse societies where the collision of horizons occurs. The first rule is an unconditional imperative: *be kind to others*. In that imperative, "kind" is to be interpreted broadly, standing in for pro-social behaviours and actions in general. This isn't merely parroting "woke" vocabulary: it just sounds more natural to say "be kind to others" than to say "engage in pro-social behaviours towards others". The second rule offers a piece of advice on how to put that imperative into practice: *don't assume everyone thinks the same way as you*. The second rule calls attention to the importance of empathy and of trying to understand how others see the world, recognising that some types of behaviour may not travel well across groups. What you think

of as kind may not be perceived the same way by someone else. What the second rule does *not* say is that a person *always* has to accommodate the wishes or desires of another. What if someone has unrealistic demands or unreasonable expectations? Negotiating the challenges presented by coexisting modern tribes will never be easy or without friction, but it helps to keep in mind that, much of the time, most people are trying to do the right thing, just in different ways.

# Notes to Part IV: Modern tribes

## 21. Joshua's question

[1] Some might wonder whether it makes sense to speak of *degrees* of irrationality. Just like a person can't be a little bit pregnant, how can a person be a little bit irrational? If one identifies *being rational* with *being consistent*, this is a fair objection: consistency is an all-or-nothing state. But being rational requires more than just being consistent. A rational person's beliefs are based on evidence and those beliefs should hang together in an overall *coherent* worldview, and coherence does come in degrees.

## 22. On polarisation

[1] There are a variety of methods used for voting in an assembly. Sometimes it is done by a voice vote, as when the chair calls out, "All those in favour, say 'Aye'…all those opposed, say 'No'." If the voice vote doesn't indicate a obvious winner, a standing vote or a show of hands may be called (which may or may not be counted). The most exact method is the recorded vote, also known as a roll-call vote, in which the name of each member along with their vote is noted.

[2] People living in Scotland, Wales, and Northern Ireland were excluded from the study, as their survey questions varied.

[3] A recent article by *The Economist* discusses how AI is increasingly being used to write news stories. One company, the British firm Radar AI, used AI to place over *400,000 news stories in local papers* since 2018, with only five people on staff. There are two concerns about this tendency. The first concern is that AI could be used to generate, *on demand*, particular news stories targeting specific issues or values for consumers, making particular things salient to them in ways that could influence their behaviour (e.g., shopping or voting). The second concern is that a world where each person receives a custom news feed is a world with fewer common reference points regarding events that shape our understanding of the world. Although you might not agree with what, for example, the *New York Times* or the *Wall Street Journal* said, old-school physical newspapers and magazines provided a fixed cultural reference point accessible to virtually everyone. If two people were talking about a subject and they had both read the same newspaper, that gave each person information about what the other knew. A world were everyone receives their news from custom AI feeds is a world where we lack common cultural reference points, and so we lack an important way of being able to get inside other peoples' heads. In addition, think about how difficult this is going to make the job of future historians (The Economist 2023).

[4] I gave ChatGPT the following prompt: "Rewrite the opening to the *Communist Manifesto* from the ideological perspective of the economist Milton Friedman." Here's what it generated:

A specter is haunting the world—the specter of free markets. All the powers of the old world have entered into a holy alliance to exorcise this specter: governments and central planners, entrenched monopolies, and the champions of bureaucratic control.
A new era of economic freedom is dawning, one that seeks to liberate individuals from the shackles of collectivism and embrace the power of voluntary exchange. We, the economists, stand at the forefront of this movement, armed with the principles of Milton Friedman, to challenge the prevailing notions of state intervention and usher in an era of market-driven prosperity.
The history of all hitherto existing societies is a history of government interference and misguided attempts to engineer economic outcomes. From planned economies to regulatory excesses, the hand of intervention has stifled innovation, undermined individual initiative, and impeded the natural forces of competition. It is time to cast aside these chains and let the invisible hand guide the destiny of nations.
Modern society is the product of countless voluntary interactions and transactions, driven by the pursuit of self-interest. It is through these actions that wealth is created, standards of living are elevated, and human progress is achieved. The free market is not a mere abstraction, but a dynamic force that harnesses the dispersed knowledge and aspirations of individuals, allocating resources efficiently and spurring innovation.
Yet, this fundamental truth is often obscured by the allure of collectivism, which promises equality but delivers mediocrity. The proponents of state control fail to recognise the inherent limitations of centralised decision-making and the dangers it poses to individual liberty. They ignore the lessons of history and disregard the perils of concentrating power in the hands of a few.
We, the advocates of free markets, reject the notion that wealth is a zero-sum game. Instead, we affirm that through voluntary exchange and mutually beneficial transactions, prosperity can be expanded and shared. It is not the accumulation of wealth by a few that impoverishes the many, but rather the stifling of economic freedom and the denial of opportunities that hinder progress.
Let the clarion call for free markets echo across the globe. Let the principles of individual liberty, limited government, and free enterprise guide our actions. Let us dismantle the barriers to competition, foster entrepreneurship, and unleash the creative energies of individuals. In doing so, we shall create a world where the pursuit of self-interest aligns with the common good, where economic liberty is the birthright of all, and where prosperity knows no boundaries.

> Working individuals of all countries, unite! Embrace the power of the free market and seize control of your economic destiny. Together, let us forge a future where the potential of every human being is unleashed, and where the promises of liberty and prosperity are not mere slogans, but a living reality for all.

## 23. Social identity, in-group bias, and norms

[1] I would like to note that my use of the phrase "modern tribe" is primarily for convenience, as it is easier to speak of a "modern tribe" than it is to speak of "a social category with which a number of people self-identify". However, there are two potential dangers in the use of this terminology, which need to be guarded against. First, the term "tribe" suggests a degree of cohesion, political unity, and solidarity of purpose, which is absent for many social categories. In anthropology, a *tribe* historically referred to a group that occupied a contiguous territory, shared a common identity, and engaged in a number of common activities, such as war, trade, and ceremonies. The modern tribes of which I speak may share some, but not necessarily all, of these features. Second, and more importantly, a number of people have objected to usage of the term "tribe" itself due to the negative connotations the term acquired during colonialism. My use of the term "tribe" is intended to be non-evaluative – a re-branding of the concept, if you like. A number of forms of social organisation that count as modern tribes are important forces for good, such as those committed to disability rights, environmental protection, animal rights, the economically disadvantaged, and so on. That said, not all modern tribes are forces for good. As Alfred noted in the film *The Dark Knight* (2008): "Some men just want to watch the world burn." The internet enables those people to organise too.

[2] For example, if a Black person is the only minority in a busy coffee shop, someone who cuts in front of them in line might well be perceived as committing a microaggression. However, if every person in the coffee shop is Black, the same line-cutting behaviour would probably not be interpreted in the same way. The set of social categories that applies to the person standing in line hasn't changed, but the set of which categories are salient, and why, has changed.

[3] Recall the discussion of social norms from Chapter 13: a modern tribe often fixes some of a person's reference network.

[4] The fact that this conflict is not felt more deeply in Western societies – both American and European – is an encouraging sign of how much our religious attitudes have evolved since the French Wars of Religion (1562–98) and the Thirty Years' War (1618–48). In those two conflicts alone, over 11 million people died.

[5] It's very unlikely that Sherif's experiments would be approved by a research ethics committee today. Nevertheless, despite their questionable ethical standing, the Sherif experiments are important to discuss because of the significance of the results and the fact that they form a building block for later work in social psychology.

[6] Some interesting cross-cultural differences exist. An extended study by Cuddy *et al.* (2009) found that in a sample of three *collectivist* cultures (Hong Kong, Japan, and Korea), the reference group for subjects tended not to be assigned high warmth and high competence. Instead, in-groups could be classified as low warmth–high competence (e.g., Japan, with the in-groups of "my university" or "students"), high warmth–low competence (e.g., Japan, with "my family" or "my friends"), or middle rankings on both dimensions. Yet even so, out-groups would continue to score low on at least one of the two dimensions and sometimes both.

## 24. The psychology of modern tribes

[1] This is essentially the same as the category of "partially anonymous" in the analysis by – seriously – Anonymous (1998): "Partial anonymity exists when either a source cannot be individually specified or when there is not a high level of knowledge about a source (but not both, which represents full anonymity)."

[2] This is a reference to a socially engineered group activity built around hatred for enemies of the state in George Orwell's dystopian novel *1984* (see Wikipedia 2024). In the novel, the main character Winston Smith describes the Two Minutes Hate as follows:

> The horrible thing about the Two Minutes Hate was not that one was obliged to act a part, but, on the contrary, that it was impossible to avoid joining in. Within thirty seconds any pretence was always unnecessary. A hideous ecstasy of fear and vindictiveness, a desire to kill, to torture, to smash faces in with a sledge-hammer, seemed to flow through the whole group of people like an electric current, turning one even against one's will into a grimacing, screaming lunatic. And yet the rage that one felt was an abstract, undirected emotion which could be switched from one object to another like the flame of a blowlamp.

It is worth noting how social media functions rather similarly, at least some of the time.

[3] At this point, its worth distinguishing the combative world of online tribal conflict and the phenomenon of *trolling*. Trolling is "the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the Internet with no apparent instrumental purpose" (Buckels *et al.* 2014, p. 97). A

number of studies have examined links between trolling and the Dark Tetrad personality traits (narcissism, Machiavellianism, psychopathy, and sadism). Buckels *et al.* (2014), in a poll of 1,215 subjects, found that the trait most heavily associated with online trolling was sadism. In a later study, Craker and March (2016), examining trolling behaviour on social networking sites, found that sadism and psychopathy were significant predictors. March *et al.* (2017) examined trolling on dating (or hook-up) apps, finding that psychopathy, sadism, and impulsivity were traits significantly associated with trolling on those apps. In all these studies, the one common trait was *sadism* – the experiencing of pleasure from causing physical or psychological pain in another person. This is interesting because online trolling is quite common, yet psychopathy and sadism are quite rare. Only a very small percentage of people, approximately 1% of the population, score high on the Hare Psychopathy Checklist. Are psychopaths and sadists just unusually busy online, or is there something about life online that draws out these characteristics from ordinary people? In one study, Pfattheicher *et al.* (2021) found a connection between boredom and sadism. We might ask, then, if people spend a lot of time on the internet because they are bored, could that be a partial explanation for why they engage in online trolling? Pfattheicher *et al.* (2021, p. 79) state explicitly that they did not investigate that, and so "it remains unclear whether boredom also relates to other online behavior, and whether boredom motivates going online in general".

[4] In the years following Allport's introduction of the contact hypothesis, much debate existed over whether it really worked. Pettigrew and Tropp (2006) conducted a meta-analysis of 515 studies, concluding that "intergroup contact typically reduces intergroup prejudice". Yet their conclusion comes with a number of caveats. For one, inter-group contact isn't a *sufficient* condition for the reduction of prejudice because, ideally, the contact takes place in an environment where the two groups have equal status and with institutional support. Thus, there are a lot of other variables that can get in the way. Furthermore, a number of studies have found that inter-group contact is not *necessary* for the reduction of prejudice.

## 25. Authenticity and the WINOs

[1] Ian Hacking discusses this phenomenon, which he refers to as an "administrative kind", in his influential book *The Social Construction of What*.

[2] Canon 1397 of the 1983 Code of Canon Law, to be precise.

[3] Although there are a lot of grains of sand in 100ml, you might guess that the number is around that of a moderately sized city – approximately two million. But since you've probably never heard of Hibbing, Minnesota, you would likely infer that it is a small town. And you would be right.

[4] In Chapter 20, I discussed the dangers resulting from the moralisation of knowledge. This is an example of dangers resulting from the moralisation of membership.

## 26. Intersectionality

## 27. Epistemic closure and extreme groups

[1] A distinct, and unrelated, sense of "epistemic closure" concerns the preservation of knowledge under entailment. For example, if a person knows some claim $p$ and also knows that $p$ entails $q$, then the person knows that $q$. Although this sense of epistemic closure is widely assumed within epistemology, in what follows, I exclusively use "epistemic closure" to refer to a system of beliefs that is closed off from the world in some way and resists being revised.

[2] Recall the discussion of falsification in the introduction.

[3] Putting matters in this way makes it sound as though epistemic closure is a *voluntary* act. Sometimes it may be but not always. Sometimes a person believes something so intensely that, when faced with evidence to the contrary, they seek to reduce cognitive dissonance by adjusting their set of beliefs without explicitly realising what they are doing.

[4] Why I have shifted from speaking of "tribes" to speaking of an "epistemic community" determined by the tribe? The point is simply this: one's tribe refers to all the members with whom you share a social identity. However, not all members of the tribe are necessarily going to be people one listens to regarding the formation of belief. For lack of a better word, certain individuals in the tribe will be earmarked as "thought leaders", who are disproportionately influential in the shaping of beliefs amongst members of the tribe. The "thought leaders" will be those individuals who are most influential, but there will often be a secondary or tertiary layer of individuals who also contribute to shaping the space of beliefs. This difference in membership is the difference between a tribe and a person's epistemic community.

[5] Furthermore, certain false beliefs can even be fitness enhancing. If a person believes that crime is on the rise and, as a result, they insist on staying home in the evenings rather than going out in public or driving places, this change in behaviour reduces their exposure to car accidents and being mugged or otherwise assaulted. This can cut both ways: Gerd Gigerenzer estimated that 1,595 Americans died in the year following 9/11 because they were so concerned about flying that they elected to drive rather than fly. Since road travel is much more dangerous than flying, this misperception of risk lead to a number of additional deaths due to road accidents (Ball 2011).

[6] This usage differs significantly from the sense intended when the term "incel" was first coined by a Canadian woman named Alana, who created a website for lonely people who considered themselves "involuntary celibates" in that they

had experienced long-term difficulty forming relationships over the course of their life.

[7] The matter, then, is analogous to the reason why the war on drugs failed despite years of effort. It is easier to try to restrict the supply of drugs than it is to try to address the social factors that generate the demand for drugs. Yet until the demand is reduced, the economic incentives ensure that new suppliers will always appear to replace those removed from the system.

## 28. The collision of horizons

[1] In what follows, I assume that all behaviours and practices I talk about are legal within the containing society. I do not necessarily assume that the behaviours and practices are *moral* because I want to allow for the possibility that the groups disagree over what counts as moral behaviour.

[2] The key issue, of course, is whether one can be said to *choose freely* when strong social norms, with sanctions attached to violations, are present. Can a person be said to choose freely to do something when the possibility of not doing that thing is associated with expulsion from the group, the loss of friends and family, and alienation from the community in which they were raised? I think not. But this is a delicate issue to think through, for nearly all choices that impact other people involve both positive and negative feedback, which can affect a person's ability to choose freely. Suppose a child must choose whether to study medicine or law or literature, knowing that their parents would be extremely happy if they chose medicine or law and extremely disappointed if they chose literature. How extreme does the parent's disappointment need to be, and how much must it affect the child's future, before we say that the child no longer faces a free choice? Even if all choices we face are "metaphysically free" choices (in the sense of free will), most of our choices have their freedom curtailed to some degree in this social sense of "free choice". Matters become more complicated when we recognise that some cultures have different moral practices that rank-order the protection of individual liberties and the protection of social norms or conventions differently. Even if *we* think, in light of a commitment to Western values and individual liberties, that it is a mistake to reverse the rank-ordering of protecting individual liberties and protecting social norms or conventions, the following question arises: to what extent does a diverse, multicultural society allow groups to make decisions – regarding their form of life, which the overarching society considers to be mistaken – before intervening? There are no easy answers to this question.

## References


Allport, Gordon W. (1954). *The Nature of Prejudice*. Addison-Wesley.

Anonymous. (1998). "To reveal or not to reveal: A theoretical model of anonymous communication." *Communication Theory* 8 (4):381–407. https://doi.org/10.1111/j.1468-2885.1998.tb00226.x

Axelrod, Robert. (December 1986). "An evolutionary approach to norms." *American Political Science Review* 80 (4):1095–1111. https://doi.org/10.1017/S0003055400185016

Ball, James. (5 September 2011). "September 11's indirect toll: road deaths linked to fearful flyers." *The Guardian*. https://www.theguardian.com/world/2011/sep/05/september-11-road-deaths

Baum, Matthew A. and Tim Groeling. (2008). "New Media and the Polarization of American Political Discourse." *Political Communication* 25:345–365. https://doi.org/10.1080/10584600802426965

Bergson, Henri. (1935). *The Two Sources of Morality and Religion*. Macmillan and Co., Limited. Translated by R. Ashley Audra and Cloudesley Brereton, with the assistance of W. Horsfall Carter.

Bicchieri, Cristina. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190622046.001.0001

Billig, Michael and Henri Tajfel. (1973). "Social categorization and similarity in intergroup behaviour." *European Journal of Social Psychology* 3 (1):27–52. https://doi.org/10.1002/ejsp.2420030103

Bleiberg, Joshua and Darrell M. West. (May 2015). "Political polarization on Facebook." Brookings Institute blog. https://www.brookings.edu/blog/techtank/2015/05/13/political-polarization-on-facebook/

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. (2017). "Greater Internet use is not associated with faster growth in political polarization among US demographic groups." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1706588114

Bramson, Aaron, Patrick Grim, Daniel J. Singer, William J. Berger, Graham Sack, Steven Fisher, Carissa Flocken, and Bennett Holman. (January 2017). "Understanding Polarization: Meanings, Measures, and Model Evaluation." *Philosophy of Science* 84 (1):115–159. https://doi.org/10.1086/688938

Buckels, Erin E., Paul D. Trapnell, and Delroy L. Paulhus. (2014). "Trolls just want to have fun." *Personality and Individual Differences* 67:97–102. https://doi.org/10.1016/j.paid.2014.01.016

Cantril, Hadley. (1941). *The Psychology of Social Movements*. New York: Wiley.
https://doi.org/10.1037/13593-000

Clark, Andy and David Chalmers. (1998). "The Extended Mind." *Analysis* 58 (1):7–19.
https://doi.org/10.1093/analys/58.1.7

Craker, Naomi and Evita March. (2016). "The dark side of Facebook©: The Dark Tetrad, negative social potency, and trolling behaviours." *Personality and Individual Differences* 102:79–84.
https://doi.org/10.1016/j.paid.2016.06.043

Crenshaw, Kimberlé. (1989). "Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics [1989]." *The University of Chicago Legal Forum* (1):139–167.

Cuddy, Amy JC, Susan T Fiske, Virginia SY Kwan, Peter Glick, Stephanie Demoulin, Jacques-Philippe Leyens, Michael Harris Bond, Jean-Claude Croizet, Naomi Ellemers, Ed Sleebos, *et al.* (2009). "Stereotype content model across cultures: Towards universal similarities and some differences." *British Journal of Social Psychology* 48 (1):1–33.
https://doi.org/10.1348/014466608X314935

Douthat, Ross. (2 May 2018). "The Redistribution of Sex." *The New York Times*.
https://www.nytimes.com/2018/05/02/opinion/incels-sex-robots-redistribution.html

Emerson, Ralph Waldo. (2005 [1841]). *Compensation and Self-Reliance*. New York: Cosimo Classics. Originally published by Fleming H. Revell Company in 1841.

Esteban, Joan-María and Debraj Ray. (July 1994). "On the Measurement of Polarization." *Econometrica* 62 (4):819.
http://doi.org/10.2307/2951734

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu. (2002). "A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition." *Journal of Personality and Social Psychology* 82 (6):878–902.
https://doi.org/10.1037/0022-3514.82.6.878

Gentzkow, Matthew and Jesse M. Shapiro. (2011). "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics* 126 (1799–1839).
https://doi.org/10.1093/qje/qjr044

Graeber, David. (2018). *Bullshit Jobs: A Theory*. Allen Lane.

Hanson, Robin. (26 April 2018). "Two Types of Envy."
http://www.overcomingbias.com/2018/04/two-types-of-envyhtml

Hare, Christopher and Keith T. Poole. (2014). "The Polarization of Contemporary American Politics." *Polity* 46 (3):411–429.
https://doi.org/10.1057/pol.2014.10

Hasler, Béatrice S. and Yair Amichai-Hamburger. (2013). "Online Intergroup Contact." In Amichai-Hamburger, Yair (ed.), *The Social Net: Understanding our online behavior*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199639540.001.0001

Hogg, Michael A. and Dominic Abrams. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*. Routledge.

Itzigsohn, José. (2004). "The Formation of Latino and Latina Panethnic Identities." In Foner, Nancy and George Frederickson (eds.), *Not Just Black and White: Historical and Contemporary Perspectives on Immigration, Race, and Ethnicity in the United States*, 197–216. New York: Russell Sage Foundation.

Jones, Stephen. (26 September 2017). "Russian 'Ultra' football hooligans unleash fury in violent footage as Liverpool and Manchester United fans fly into Moscow." *The Mirror*. https://www.mirror.co.uk/news/world-news/russian-ultra-football-hooligans-unleash-11241338

Kimmel, Michael. (2018). *Healing from Hate: How Young Men Get Into — and out of — Violent Extremism*. University of California Press. https://doi.org/10.1525/9780520966086

Lapidot-Lefler, Noam and Azy Barak. (2012). "Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition." *Computers in Human Behavior* 28:434–443. https://doi.org/10.1016/j.chb.2011.10.014

Lewis, Helen. (20 February 2014). "The uses and abuses of intersectionality." *New Statesman*. https://www.newstatesman.com/politics/2014/02/uses-and-abuses-intersectionality

List, Christian and Philip Pettit. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199591565.001.0001

Malinowski, Bransław. (1941). *A Scientific Theory of Culture, and other essays*. Chapel Hill: University of North Carolina Press.

March, Evita, Rachel Grieve, Jessica Marrington, and Peter K. Jonason. (2017). "Trolling on Tinder© (and other dating apps): Examining the role of the Dark Tetrad and impulsivity." *Personality and Individual Differences* 110:139–143. https://doi.org/10.1016/j.paid.2017.01.025

Mencken, Henry Louis. (1982). *A Mencken Chrestomathy*. Vintage Books. Original published: New York, Knopf, 1949.

Milanovic, Branko. (2012). "Global Income Inequality by the Numbers: in History and Now." Technical report, The World Bank, Development Research Group. Policy Research Working Paper 6259. https://doi.org/10.1596/1813-9450-6259

Nydegger, R. V. and G. Owen. (1974). "Two-Person Bargaining: An Experimental Test of the Nash Axioms." *International Journal of Game*

*Theory* 3 (4):239–249.
https://doi.org/10.1007/BF01766877

Parekh, Bhikhu. (2008). *A New Politics of Identity: Political Principles for an Interdependent World.* Palgrave Macmillan.

Pettigrew, Thomas F. and Linda R. Tropp. (2006). "A Meta-Analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology* 90 (5):751–783.
https://doi.org/10.1037/0022-3514.90.5.751

Pettit, Philip. (1995). "The Virtual Reality of *Homo Economicus*." *Monist* 78 (3):308–329.
https://doi.org/10.5840/monist199578322

Pew Research Center. (4 March 2015). "Most Say Government Policies Since Recession Have Done Little to Help Middle Class, Poor."
https://www.pewresearch.org/politics/2015/03/04/most-say-government-policies-since-recession-have-done-little-to-help-middle-class-poor/

Pfattheicher, Stefan, Ljiljana B. Lazarević, Erin C. Westgate, and Simon Schindler. (September 2021). "On the relation of boredom and sadistic aggression." *Journal of Personality and Social Psychology* 121 (3):573–600.
http://dx.doi.org/10.1037/pspi0000335

Pinker, Steven. (2011). *The Better Angels of our Nature: Why Violence has Declined.* Viking.

Popper, Karl R. (1945). *The Open Society and Its Enemies: Volume One — The Spell of Plato.* Routledge & Kegan Paul.

Proctor, Bernadette D., Jessica L. Semega, and Melissa A. Kollar. (2016). "Income and Poverty in the United States: 2015." Technical report, United States Census Bureau.

Radcliffe-Brown, A. R. (1952). *Structure and Function in Primitive Society.* London: Cohen and West.

Ruesch, Michelle. (2013). "A Peaceful Net? Intergroup Contact and Communicative Conflict Resolution of the Israel-Palestine Conflict on Facebook." In Ternès, Anabel (ed.), *Communication: Breakdowns and Breakthroughs*. Oxford, United Kingdom: Inter-Disciplinary Press.

Sherif, Muzafer. (ed.). (1962). *Intergroup Relations and Leadership.* New York: Wiley.

Sherif, Muzafer. (1966). *In Common Predicament: Social Psychology of Intergroup Conflict and Cooperation.* Boston: Houghton-Mifflin.

Sherif, Muzafer and Carolyn W. Sherif. (1953). *Groups in Harmony and Tension.* Octagon Books.

Sinclair, Upton. (1995 [1934]). *I, Candidate for Governor, and How I Got Licked.* University of California Press. Originally published: New York: Farrar & Rinehart, © 1935.

Srinivasan, Amia. (2018). "Does anyone have the right to sex?" *London Review of Books* 40 (6):5–10.

Suler, John. (2004). "The Online Disinhibition Effect." *CyberPsychology & Behavior* 7 (3).
https://doi.org/10.1089/1094931041291295

Sunstein, Cass R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Tajfel, Henri. (1972). "Social categorization." In Moscovici, S. (ed.), *Introduction à la psychologie sociale*, volume 1. Paris: Larousse.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. (1971). "Social categorization and intergroup behaviour." *European Journal of Social Psychology* 1 (2):149–178.
https://doi.org/10.1002/ejsp.2420010202

The Economist. (4 May 2023). "Artificial intelligence is remixing journalism into a "soup" of language." *The Economist*.
https://www.economist.com/business/2023/05/04/artificial-intelligence-is-remixing-journalism-into-a-soup-of-language

Turner, J. C. (1982). "Towards a cognitive redefinition of the social group." In Tajfel, H. (ed.), *Social Identity and Intergroup Relations*. Cambridge University Press.

Wheatley, Jonathan. (2015). "Restructuring the policy space in England: The end of the Left—Right paradigm?" *British Politics* 10 (3):268–285.
https://doi.org/10.1057/bp.2015.35

Wheatley, Jonathan. (6 June 2017). "The polarisation of party supporters since 2015 and the problem of the 'empty centre' — in maps." LSE British Politics and Policy blog.
https://blogs.lse.ac.uk/politicsandpolicy/the-polarisation-of-party-supporters-since-2015/

Wheaton, Sarah. (18 June 2007). "Political Sites Drawing More Eyeballs." *The New York Times*.
https://thecaucus.blogs.nytimes.com/2007/06/18/political-sites-drawing-more-eyeballs/

Wikipedia. (2024). "Two Minutes Hate."
https://en.wikipedia.org/wiki/Two_Minutes_Hate