THE ECONOMIC JOURNAL

The Economic Journal, **134** (*April*), 1019–1068 https://doi.org/10.1093/ej/uead094
Advance Access Publication Date: 26 October 2023

RADICALISATION*

Jean-Paul Carvalho and Michael Sacks

This paper analyses the rise of radical movements and the design of counter-radicalisation policies. A group derives meaning from participation in identity-based activities and a forward-looking organisation provides a platform for these activities. The warning sign for radicalisation is cultural purification by the organisation, i.e., the screening out of moderates and exclusive recruitment of radicals. While this shrinks the club, it puts it on a growth path along which it becomes larger and more extreme over time. Conventional counter-radicalisation policies can backfire. The radicalisation mechanisms we identify can be disabled by mild anti-radical messaging and informational interventions that eliminate stereotypes.

Political and economic development can be destabilised by the formation of radical religious, ethnic and ideological movements. Perhaps the most striking example is the rise of Nazism in Weimar Germany, a radical ideological cult that culturally transformed the country, and seized control of the state (Burleigh, 2000). After the Cold War, security interests turned to the risk posed by radical non-state actors, both domestic and foreign. Much of the focus was on the social integration and radicalisation of Muslims in the West, as well as Islamist militant groups operating in the Middle East, North Africa and Asia (e.g., Kepel, 2003; Roy, 2017). Other examples of radical movements include doomsday cults in the United States and Japan (Reader, 2013; Wright, 2014), separatist groups such as the IRA, LTTE and ETA, and far-left guerrillas such as Italy's Red Brigades (Sundquist, 2010). A more recent concern is the populist and ethnonationalist movements taking root in the United States and Europe (Fukuyama, 2018; Dal Bo *et al*., 2023).

This paper presents a theory of radicalisation. Combining a club model from the economics of religion (Iannaccone, 1992; Berman, 2000; Levy and Razin, 2012) with intergenerational cultural transmission (Bisin and Verdier, 2000; 2001), we identify three mechanisms driving radical movements and explore their implications for the design and implementation of counter-radicalisation policies. The model is inspired by a common theme of radical movements, the

* Corresponding author: Jean-Paul Carvalho, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK. Email: jean-paul.carvalho@economics.ox.ac.uk

This paper was received on 8 March 2021 and accepted on 20 October 2023. The Editor was Steffen Huck.

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.8383871.

We are grateful for comments by the editor and five anonymous referees, as well as Alberto Alesina, Ran Abramitzky, Robert Barro, Samuel Bazzi, Eli Berman, Lisa Blaydes, Gary Cox, Martin Fiszbein, Larry Iannaccone, Saumitra Jha, Patrick Julius, Asim Khwaja, Timur Kuran, David Laitin, Mike McBride, Rachel McCleary, Alireza Naghavi, Nathan Nunn, Debraj Ray, Jared Rubin, Moses Shayo, Stergios Skaperdas, Angela Vossmeyer, Cody Weller, Peyton Young, participants at the 2023 AALIMS Conference, 2018 ASREC conference, 2017 ERINN Conference, 2017 AALIMS Conference, 2017 NES Conference 'Towards Effective and Equitable Development: The Role of Institutions and Diversity', 2016 ASREC conference, 2015 IMBS Workshop on the Economics of Religion and Culture, and seminar participants at Oxford, Harvard, Stanford, Boston University, Manchester and UC Riverside. Aniket Chakravorty and Anna Ellithorpe provided excellent research assistance. A previous working paper, circulated under the title 'The Economics of Religious Communities: Social Integration, Discrimination and Radicalisation', was split into this and a separate paper titled 'The Economics of Religious Communities'.

exploitation of latent but powerful forms of identity (Berger, 2018).[1] Identity-based organisations exploit a demand for meaning and belonging by members of an identity group (e.g., religion, ethnic group, social class) by coordinating participation in rituals, ideology and activism. In doing so, they can transform society both culturally and politically (Carvalho, 2016).

To illustrate the role of identity in radical movements, consider the Muslim Brotherhood, the progenitor of the various Islamist movements of the twentieth century (Mitchell, 1993; Wickham, 2015). The Brotherhood was founded in 1928 by Egyptian schoolteacher Hassan al-Banna in the minor city of Ismailia. It was a fringe group promoting a pan-Islamic identity after decades of colonial dominance and secularisation. Originally concentrating on education and social service, it eventually built a state within a state through a network of highly committed lay volunteers:

> In the 1940s, the Egyptian Muslim Brotherhood developed a network of schools, clinics, and even small industries. They created student and professional sections, geared in fact toward modern professions (lawyers, engineers, doctors, teachers, bureaucrats), as well as worker sections in which they encouraged unionisation.
>
> (Roy, 1994, p. 57)

The Brotherhood's aim was to '[...] become those whom governments task with preparing the curricula and selecting teachers for Islamic education in public schools, appointing imams in public institutions such as the military, the police, or prisons, and receiving subsidies to administer various social services' (Vidino, 2020, p. 10). Through this religious ecology, the Brotherhood 'resocialised' society along Islamic lines and ultimately seized political power. Its approach was replicated across the Middle East and in the West, where the goal was 'the fostering of a strong, resilient, and assertive Islamic identity among Western Muslims' (Vidino, 2020, p. 9). The success of the Muslim Brotherhood and its offshoots was evident in more visible forms of piety, including Islamic banking and the new veiling movement (Kuran, 2004; Carvalho, 2013; Binzel and Carvalho, 2017), as well as growing political influence (Kepel, 2003; Wickham, 2015). From these organisations also came militant groups that pursued political power through violent means (Berman, 2009). For example, Takfir wal-Hijra spun off from the Muslim Brotherhood in 1971, attempted to topple the Egyptian state through terrorist attacks and produced Ayman Al-Zawahiri, later leader of Al-Qaeda (Ibrahim, 1980).

To analyse the radicalisation of an identity group, we develop an infinite-horizon model in which overlapping generations participate in identity-based activities. The identity group could be an entire population or a religious, ethnic or linguistic group; a subpopulation with shared ideas or experiences; or a sociodemographic class. A long-lived organisation provides the platform for identity-based activities, such as rituals, ideological immersion, social interaction, political activism and other forms of collective investment in identity. We refer to this platform as 'the club', and those who join it as 'club members'. Each period, the organisation sets a participation requirement (strictness) for club membership and group members decide whether to join or be unaffiliated. Club members gain access to a collectively produced club good, which we interpret as a sense of meaning and belonging, but which could also represent more tangible benefits from group activism (e.g., political rents). While members of the group have a common identity, they differ in their degree of identification, being either moderate or radical. Radical types place greater

[1] The importance of identity to economics was established by Akerlof and Kranton (2000; 2010), as well as Chen and Li (2009), Shayo (2009), Bénabou and Tirole (2011) and Akerlof (2017).

value on the club good, and thus prefer more extreme levels of identity-based participation.[2] Both types are culturally intolerant toward each other: radicals prefer that their children acquire the radical trait, whereas moderates prefer that their children acquire the moderate trait.

The share of radicals in the identity group evolves through a process of intergenerational cultural transmission, consisting of three parts: (*i*) club-level transmission, (*ii*) institutional transmission and (*iii*) group-level transmission. Club-level transmission occurs when a child is matched uniformly at random with a club member from their parent's generation and acquires their trait (i.e., moderate or radical). The likelihood of this occurring is the share of time their parent spends in the club. While this is a natural form of cultural transmission in endogenous communities, it has received little attention in prior work.[3] Institutional transmission occurs when a child is exposed to a central socialising institution (e.g., state media, public education) and acquires the moderate trait.[4] Thus, the state's objective is to moderate the identity group, and a radical club can be considered an anti-establishment movement. Group-level transmission is the familiar oblique transmission channel in which a child is matched uniformly at random with a group member from their parent's generation and acquires their trait (Cavalli-Sforza and Feldman, 1981). Through club-level transmission, current participation in the club affects the future share of radicals in the identity group.

We deem that radicalisation occurs when the share of radicals is monotonically increasing over time. Radicalism occurs when the share of radicals is stabilised above zero. Otherwise, there is moderation. In setting strictness, we assume that the organisation's objective is to maximise a discounted sum of participation in the club over time, not to radicalise per se. As identity-based participation affects cultural transmission, an organisation can face a trade-off between maximising current participation and raising the share of radicals in order to boost future participation. As we shall see, there are conditions under which the organisation's payoff is maximised by tuning strictness over time in a manner that produces radicalisation. Under other conditions, its payoff is maximised by inducing moderation.

We identify three mechanisms that lead to radicalisation. First, radicalisation can occur through cultural purification of the club, i.e., when the organisation screens out moderates and caters only to radicals at high strictness. This boosts the share of radicals both by restricting transmission within the club to the radical trait and by increasing the time spent in the club, thereby shielding radicals from institutional transmission. Through group-level transmission, this vanguard of radicals converts moderates and radicalises the identity group over time. Because of the screening out of moderates, participation in identity-based activities might decline in the short term and the organisation appear to weaken. However, cultural purification puts the club on a growth path along which it becomes larger and more extreme over time. For the organisation to be willing and able to radicalise the group in this manner, we show that cultural intolerance must be sufficiently high. Typically, there are also multiple steady states, with radicalisation only occurring when the initial share of radicals is large enough.

Two additional mechanisms fuel this process of radicalisation. Suppose that active participants (club members) are seen as role models, due, for example, to greater prestige, visibility

[2] This is the standard definition of radicals in the economics of religion literature (see Berman, 2009). Violence is not required for radicalism, though we believe that our theory applies to both non-violent and violent radical movements.

[3] This type of transmission is similar to the treatment of homogamy in Bisin and Verdier (2000). The difference here is that the distribution of types in the 'restricted pool', i.e., the club, is endogenous.

[4] By accounting for institutional transmission, we follow Carvalho (2016), Carvalho *et al.* (2017; 2022) and Verdier and Zenou (2018).

or outreach activities. This might confer on club members disproportionate cultural influence in group-level transmission. We call this biased cultural transmission. In this case, radicalisation can occur irrespective of the initial share of radicals. In addition, if extreme levels of identity-based participation cause the broader society to discriminate against group members then raising strictness reduces outside options for all members of the identity group. This allows the organisation to become even stricter, further reducing the outside options of group members, and so forth. In this way, the organisation can carve out a niche, shielding itself from outside competitive pressures. We remark that this process can be activated by rising fears of extremism, which then become self-fulfilling. These two mechanisms are related to important concepts in evolutionary theory, namely, prestige-biased cultural transmission (Henrich and Gil-White, 2001) and niche construction (Odling-Smee *et al*., 2003), both of which have been largely ignored in economics.[5]

Our analysis has important and surprising implications for counter-radicalisation policy. It reveals why conventional policies might fail, by activating one or more of the three mechanisms we have identified. For reasons we set out, stigmatising participation in identity-based activities, 'hearts and minds' strategies for winning over moderates, making moderates more intolerant of radicals and strengthening institutional transmission of the moderate trait can all backfire and boost radicalisation. More broadly, assimilationist policies, such as French secularism (*laïcité*), can promote radicalisation by making it impossible to form an inclusive club at low strictness (as moderate types exit). This leaves an identity-based organisation no choice but to form a strict club composed of radicals, who eventually radicalise the identity group. Thus, assimilationist policies can create a synthetic radicalisation strategy.

Finally, our theory suggests three principles for counter-radicalisation policy. First, a policymaker must unite and moderate: if it cannot completely eliminate identity-based participation, it should unite the identity group in an inclusive club. Any move that breaks up an inclusive club will promote radicalisation through the three mechanisms identified. Second, the policymaker must take a middle path in its anti-radical messaging. If transmission through socialising institutions (e.g., state media, public education) is too weak, institutional transmission cannot offset the effect of an exclusive club in fostering radicalism. If institutional transmission is too strong, however, club membership becomes so valuable to radicals as a shield from institutional transmission that the organisation can overwhelm institutional transmission by raising strictness and using radical club members to radicalise the identity group. Third, the policymaker must eliminate stereotypes that equate all members of the group with its most extreme participants. Internally, this would reduce the prestige and cultural influence of active club members and thereby weaken bias in cultural transmission. Externally, it would reduce blanket discrimination against group members and thereby inhibit niche construction. To this end, one concrete policy is for governments in Europe and elsewhere to abandon their strategy of appointing religious leaders as representatives of diverse groups (including secular individuals). Thus, our analysis accords with Sen (2006), who attributes violent radicalism to the 'imposition of singular and belligerent identities' and their exploitation by 'artisans of terror' (p. 2).

Our theory of radicalisation is based on a dynamic club model that builds on the static club model of Carvalho and Sacks (2021). We add to the static model overlapping generations, intergenerational cultural transmission, parental concern for their child's identity and dynamic strategies by organisations. The three radicalisation mechanisms we identify and all insights into

[5] Notable exceptions are Han *et al.* (2021) on self-enhancing bias in transmission of investment returns and Carvalho and Koyama (2016, Section 5.2) on niche construction by religious sects.

counter-radicalisation policy come from the dynamic analysis here.[6] The most closely related work to ours is as follows. In their study of Jewish emancipation, Carvalho and Koyama (2016) also combined concepts from religious club and cultural transmission models (Section 4.1). Two important papers take this substantially further: Verdier and Zenou (2015) analysed cultural contests between competing (myopic) leaders pushing two different cultural traits, while Verdier and Zenou (2018) combined club goods production and cultural transmission, introducing new dynamic techniques to model forward-looking leaders. In their simulation-based analysis, Fan *et al.* (2021) found that religious fundamentalism is associated with higher real income, lower social capital and lower compatibility between religious and secular consumption. On the role of leaders in cultural transmission, see also Hauk and Mueller (2015), Bisin and Verdier (2017), Prummer and Siedlarek (2017), Chen *et al.* (2019), Almagro and Andrés-Cerezo (2020) and Iyigun *et al.* (2021). None of the three mechanisms we identify, and hence none of our insights into counter-radicalisation policy, is featured in this work.[7]

The remainder of the paper is structured as follows. Section 1 sets out the baseline model of dynamic club formation. Section 2 analyses radicalisation through cultural purification, while Section 3 shows how this mechanism causes existing counter-radicalisation policies to fail. Section 4 introduces and analyses the two additional mechanisms, biased cultural transmission and endogenous discrimination. Tying together the analysis, Section 5 evaluates existing approaches to cultural diversity and sets out new principles for preventing radicalisation. To demonstrate the theory's relevance, our discussion is grounded in the case of European Muslims. Section 6 concludes.

1. The Dynamic Club Model

Time is discrete and indexed by $t = 0, 1, 2, \ldots$. At each time t, there is a set of individuals I known as the *identity group*, which is a continuum of players: $I = [0, 1]$, endowed with Lebesgue measure λ and Borel σ-algebra $\mathcal{B}$. Each member of the identity group $i \in I$ is born with a type $\tau_i \in \{m, r\}$, where m is a moderate type and r is a radical type. There is also an organisation that provides a platform, i.e., club, for identity-based activities. Individuals choose whether to join the club or be unaffiliated. Let $x_i^t \in \{0, 1\}$ denote i's affiliation decision at time t, where $x_i^t = 1$ corresponds to affiliated at time t and $x_i^t = 0$ corresponds to unaffiliated at time t. The organisation sets a participation requirement s^t: at each time t, every member of the club must devote share s^t of their time to the club's identity-based activities. This participation requirement is dynamically tuned by the organisation to maximise total participation in the club, taking into account the effect of current levels of participation on the future distribution of types.

Timing. Each member of I in generation t lives for one period and asexually produces a child. Interactions in each period take place as follows.

(*1*) Date 0. The set of individuals I is partitioned into two measurable subsets of players, moderates $I_m^t = (p^t, 1]$ and radicals $I_r^t = [0, p^t]$. The share of radicals, p^t, is endogenously determined from $t = 1$ onward, with initial condition $p^0 \in (0, 1)$.

[6] Other work on the religious club model includes Iannaccone and Berman (2006), McBride (2008; 2010; 2015), Aimone *et al.* (2013), Levy and Razin (2014), Carvalho (2016; 2020) and Carvalho and Koyama (2016). See Carvalho (2019) for a review.

[7] For example, Verdier and Zenou (2018) assumed that members of a given group are homogeneous, so the cultural purification mechanism, in which the organisation uses a vanguard of radicals to radicalise moderates, does not operate.

(*2*) DATE 1. The organisation announces its participation requirement $s^t \in [0, 1]$. We refer to s^t as the organisation's *strictness*.

(*3*) DATE 2. Observing strictness s^t, individuals choose whether to join the club or be unaffiliated. The set of club members at time t is denoted by $M^t = \{i : x_i^t = 1\}$. Each individual then divides one unit of time between production of a club good (i.e., participation in identity-based activities) and work/leisure outside the club. Each individual spends $x_i^t s^t$ units of time on participation in the club.

(*4*) DATE 3. A club good is collectively produced and children are socialised as follows.

  (*i*) *Club-level transmission.* With probability s^t—the share of time the parent spends on club participation—the child of a club member is matched with someone from her parent's club uniformly at random and acquires their trait. Denote the (endogenous) share of radical types in the club by q^t. Then, if club-level transmission occurs, the child becomes a radical type r with probability q^t and a moderate type m with probability $1 - q^t$.

  With complementary probability $1 - s^t$, club-level transmission fails and the child's type is determined by either institutional or group-level transmission. Note that club-level transmission fails with probability one for children of unaffiliated parents.

  (*ii*) *Institutional transmission.* If club-level transmission fails then, with probability $\gamma \in (0, 1)$, the child's type is determined by transmission from a socialising institution (e.g., state media, public education), in which case they become a moderate type m.

  (*iii*) *Group-level transmission.* Otherwise, with probability $1 - \gamma$, the child's type is determined by uniform random matching with a member of generation t: with probability p^t, the child becomes a radical and, with probability $1 - p^t$, the child becomes a moderate.

The cultural transmission process is illustrated by Figure 1.[8]

Hence, the probability that a club member becomes a τ' type is

$$\rho_{\tau'}^t = \begin{cases} s^t(1 - q^t) + (1 - s^t)[\gamma + (1 - \gamma)(1 - p^t)] & \text{if } \tau' = m, \\ s^t q^t + (1 - s^t)(1 - \gamma)p^t & \text{if } \tau' = r. \end{cases}$$

The probability that an unaffiliated individual acquires trait τ' is the same except with s^t set to zero.

At the end of period t, generation t dies and the population shares update. The evolutionary dynamic arising from the cultural transmission process (*i*)–(*iii*) above is

$$p^{t+1} = \lambda(M^t)[s^t q^t + (1 - s^t)(1 - \gamma)p^t] + (1 - \lambda(M^t))(1 - \gamma)p^t, \tag{1}$$

where $\lambda(M^t)$ is the measure of the set of members of the club at time t, M^t.

Organisation payoffs. Unlike individuals, the organisation has an infinite horizon and maximises participation in its club over time. Let us now write the set of club members at t as a function of strictness s^t and the state p^t: $M^t(s^t, p^t)$. Total participation at time t is denoted by $X^t(s^t, p^t) := \int_I x_i^t s^t d\lambda = \lambda(M^t(s^t, p^t))s^t$. Let $\mu \in [0, 1)$ be the organisation's discount factor. We assume that in every period T, the organisation chooses s^t to maximise the following

[8] We ignore vertical transmission from parent to child, instead pursuing the other approach to cultural transmission set out by Bisin and Verdier (2000), which is relatively understudied. That is, parents not only directly teach children, but also join groups/clubs in which children are socialised. The difference in our model is that the composition of the club q is endogenous and central to the analysis.

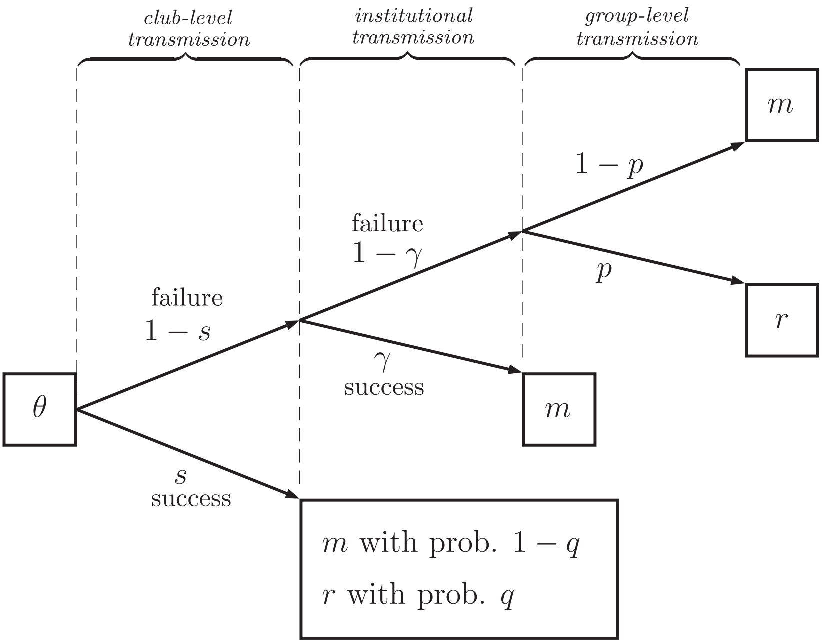

Fig. 1. *Cultural Transmission Probabilities for the Child of a Type-τ Parent Who Joins the Club.*
Notes: Here s is the club's strictness, q is the share of radicals in the club and γ is the strength of institutional transmission. Cultural transmission probabilities for an unaffiliated individual are the same, except with $s = 0$.

discounted sum of total participation:

$$G^T = \sum_{t=T}^{\infty} \mu^{t-T} X^t(s^t, p^t) \tag{2}$$

subject to the cultural dynamic (1).

Individual payoffs. Each individual's payoff has three components, their payoff from private consumption, identity-based activities (e.g., social interaction, rituals) and their child's type. We model identity-based activities as producing a club good that can be interpreted as a collectively produced sense of meaning or belonging (Carvalho, forthcoming). Importantly, radical types value this club good more than moderates. In addition, each parent has preferences over their offspring's type, preferring their child to have the same type as their own. We normalise the payoff to a τ type from having a child of type $\tau' \neq \tau$ to zero. Their payoff from a type τ child is denoted by Δ, which we assume is the same for each type. The standard interpretation of Δ is that it measures the degree of cultural intolerance, i.e., how much a person dislikes their child being of a different type.

As each generation lives for one period, we can drop time notation for the moment. Let $x(s, p) = (x_i(s, p))_{i \in I}$ be the participation strategy profile in state p given strictness s. The payoff to a club member i ($x_i(s, p) = 1$) of type τ is then

$$u_i(x(s, p); s) = \underbrace{\pi(1-s)}_{\text{private consumption}} + \underbrace{\theta_\tau X^{1/2}}_{\text{club good}} + \underbrace{\rho_\tau \Delta}_{\text{child's type}}, \tag{3}$$

where $\pi > 0$ is the payoff per unit of time spent outside of the group (e.g., real wage) and θ_τ is the value placed by type τ on the club good produced. We assume that $0 < \theta_m < \theta_r$, so that both types value the club good, but radical types do so more than moderate types. Hence, radicals are individuals who derive greater meaning and belonging from identity-based participation and are thereby willing to engage in more extreme levels of participation. Again, X is total participation in the club, ρ_τ is the likelihood that a club member has a child of type τ and Δ is the degree of cultural intolerance.

As non-members are excluded from club participation and consumption of the club good, the payoff to an unaffiliated player of type τ is $\pi + \rho_\tau \Delta$, where $s = 0$ in computing ρ_τ.

To focus on interior solutions and avoid uninteresting boundary conditions, we assume that $\pi > \theta_m$ and $\pi > \Delta$.[9] The structure of the game is common knowledge.

Equilibrium definition. Our characterisation of radicalisation can be broken into two parts. First, we consider the membership choices by each generation, accounting for coalitional deviations that are central to the formation and fragmentation of clubs. Second, we characterise the strictness strategy of the organisation given the membership responses of group members.

In each state p, and after every possible choice of strictness s, each player $i \in I$ must decide whether to join the club or be unaffiliated. We call this the (s, p)-*membership game*. Our solution concept accounts for deviations by coalitions belonging to the set $\mathcal{C} = \mathcal{B} \setminus \{\emptyset\}$, i.e., the non-empty measurable subsets of the player set. In particular, we focus on the strong Nash equilibrium (Aumann, 1959) of the (s, p)-membership game where coalitional deviations are restricted to coalitions $C \in \mathcal{C} = \mathcal{B} \setminus \{\emptyset\}$.

The coalition proofness of a strong Nash equilibrium rules out *coordination failures* based on self-fulfilling expectations that Pareto-dominated configurations of club membership will arise.

To make for cleaner statements, best-response ties are broken in the following manner.

(*I*) Whenever indifferent, a coalition joins the largest club its members are indifferent between joining.

(*II*) Whenever the organisation is indifferent between strictness levels, it chooses the one that implements the largest club.

For the entire game, our equilibrium definition combines Markov strategies for the organisation with the strong Nash equilibrium in each (s, p)-membership game.

DEFINITION 1. *An identity equilibrium (IE) of the entire game consists of*

(*i*) *participation strategies that form a strong Nash equilibrium $x^*(s, p)$ for each $(s, p) \in [0, 1]^2$ and respect tie-breaking rule (I), and*

(*ii*) *a function $s^*(p)$ such that in each state p, strictness $s^*(p)$ maximises (2) and respects tie-breaking rule (II).*

That is, an identity equilibrium is a Markov perfect equilibrium that allows clubs to form and fragment through coalitional deviations.

Dynamic definitions. Equilibrium dynamics are characterised in the following terms.

[9] When $\theta_m \geq \pi$, the organisation always forms a club composed of all group members I at full strictness $s = 1$ and the share of moderates converges to one. Similarly, when $\pi \leq \Delta$, the organisation can always form a radical organisation at full strictness $s = 1$, even when p is arbitrarily small. The dynamics then proceed as described when full strictness is chosen.

DEFINITION 2. *The dynamic exhibits radicalism if there is always a positive share of radicals: $p^t > 0$ for all t. Otherwise, we say that the dynamic exhibits moderation.*

Whereas radicalism and moderation have to do with the limiting behaviour of the cultural dynamic, radicalisation is a characteristic of the transition path.

DEFINITION 3. *Radicalisation occurs if there exists a finite time T such that the share of radicals is increasing for all $t \geq T$: $(p^t)_{t=T}^{\infty}$ is an increasing sequence. Total radicalisation occurs if, in addition, $\lim_{t\to\infty} p^t = 1$.*[10]

We describe the dynamics in terms of the share of radicals, as p is the natural state variable in this context and because strictness and total participation typically track the share of radicals, with some exceptions that we point out below.

2. Baseline Radicalisation

We now study our first mechanism for radicalisation, which occurs through the cultural purification of the club. The two further mechanisms introduced in Section 4 piggyback on this mechanism.

2.1. *Membership and Strictness Choices*

The payoff to i from joining a club is increasing in the mass of type τ_i individuals who join (see (3)). This means that either all type τ individuals join a club or none do. Hence, we need only consider type-symmetric deviations by coalitions $\mathcal{C} \in \{I_m, I_r, I\}$. Coalitional deviations have two consequences. First, because the payoff to joining is no less for radicals than moderates, if all moderates join, so do all radicals, i.e., $M^*(s, p) \in \{\emptyset, I_r, I\}$. Second, $M(s, p) = I$ is implemented by an IE if non-affiliation yields a lower payoff to moderates than joining the club when all other players join. Membership $M(s, p) = I_r$ is implemented by an IE if non-affiliation yields a lower payoff to radicals and a higher payoff to moderates than joining the club when all radicals (only) join. Otherwise, all players choose non-affiliation. We refer to a club that all identity group members join, $M^*(s, p) = I$, as *inclusive*. A club that only radicals join, $M^*(s) = I_r$, is *exclusive*.

The organisation can always form an inclusive club: for all $p \in [0, 1]$, there exists an $s > 0$ such that $M^*(s, p) = I$ (Lemma 3 in the Appendix). Hence, the organisation will never set s such that $M^*(s, p) = \emptyset$, as this would lead to a strictly lower payoff. However, even if the organisation can form an inclusive club at some level of strictness, it might instead maximise total participation by setting a higher level of strictness, *screening* out moderate types, and extracting larger contributions from radical types. Whereas the screening of moderates is simply assumed in standard religious club models, screening here emerges endogenously.

As the organisation wishes to maximise participation, it will set strictness as high as possible. For an inclusive club, that means setting strictness up to the maximum level that satisfies the moderate type's participation constraint, IR_m:

$$\underbrace{\pi(1-s) + \theta_m(s)^{1/2} + [s(1-p) + (1-s)(\gamma + (1-\gamma)(1-p))]\Delta}_{\text{moderate's payoff in an inclusive club}} \geq \underbrace{\pi + [\gamma + (1-\gamma)(1-p)]\Delta}_{\text{moderate's unaffiliated payoff}}.$$

[10] Since this sequence is monotone and bounded, asymptotic convergence is guaranteed by the monotone convergence theorem.

The maximum strictness satisfying IR_m is given by

$$\underline{s}(p) = \left(\frac{\theta_m}{\pi + \gamma p \Delta}\right)^2. \tag{4}$$

Note that $\underline{s}(p) \in (0, 1)$ because $0 < \theta_m < \pi$ by assumption. When myopic, the organisation sets strictness $\underline{s}(p)$ when forming an inclusive club. When the organisation is forward looking, $\underline{s}(p)$ may no longer be precisely the strictness of an inclusive club, but rather the *maximal* strictness that can be set for an inclusive club. It could be that the organisation forms an inclusive club at lower strictness to reduce the share of radicals in the group at a faster rate. Note, however, that in the limit as $p^t \to 0$, a (homogeneous) inclusive club will still form at strictness $\underline{s}(0)$.

For an exclusive club, participation is maximised by setting strictness up to $s = 1$ or the maximum strictness that satisfies the radical type's participation constraint, IR_r:

$$\underbrace{\pi(1-s) + \theta_r\,(ps)^{1/2} + [s + (1-s)(1-\gamma)p]\Delta}_{\text{radical's payoff in an exclusive club}} \geq \underbrace{\pi + (1-\gamma)p\Delta.}_{\text{radical's unaffiliated payoff}}$$

The maximum strictness satisfying IR_r is given by

$$\bar{s}(p) = \min\left\{\left(\frac{\sqrt{p}\theta_r}{\pi - \Delta + (1-\gamma)p\Delta}\right)^2, 1\right\}. \tag{5}$$

There are two countervailing effects of the share of radicals p on the strictness of an exclusive club. First, due to increasing returns to club goods production, the payoff to a radical from joining an exclusive club is increasing in the number of other radicals that join p. On the other hand, group-level transmission is less risky for radicals when p is large, making the protection provided by the club from group-level transmission less valuable (see Figure 1).[11] These countervailing effects mean that strictness $\bar{s}(p)$ can be a non-monotonic function of p, at first increasing in p and then decreasing. As such, it can be that $\bar{s}(p)$ equals one for intermediate values of p and less than one when the share of radicals is either small or large.

Another remarkable feature is that the strictness of an inclusive club $\underline{s}(p)$ is decreasing in cultural intolerance Δ, whereas the strictness of an exclusive club $\bar{s}(p)$ is increasing in Δ (ignoring boundary conditions). The reason is that club membership (partially) replaces institutional transmission of the moderate trait with exposure to radicals within the club (see Figure 1). This is costly to moderates. The greater the cultural intolerance Δ, the greater this cost and the lower $\underline{s}(p)$ must be to keep moderates in the inclusive club. In contrast, the substitution of institutional transmission with exposure to radicals within the club is a benefit to radicals. The greater the cultural intolerance Δ, the greater this benefit and the higher the maximum strictness $\bar{s}(p)$ that keeps radicals in the exclusive club.

2.2. *Dynamics Conditional on Membership and Strictness*

Before diving into the complete dynamics, we first illustrate how the distribution of moderates and radicals evolves over time given each of the two types of club the organisation is able to form: an inclusive club or exclusive club. We first show that an inclusive club acts as a moderating force, implying that exclusivity is necessary for radicalisation.

[11] This is the strategic substitutability introduced by Bisin and Verdier (2000; 2001), which preserves cultural diversity in their model. In our model, the difference is that club-level transmission occurs within endogenously formed clubs rather than households. The effect is to keep the strictness of an exclusive club interior at high levels of p.

Table 1. *Threshold Values for Radicalism.*

$$\hat{p} = \sqrt{\left(\frac{\theta_r \pi - (1-\gamma)\theta_m \Delta}{2\gamma\theta_r \Delta}\right)^2 + \frac{\theta_m(\pi - \Delta)}{\gamma\theta_r\Delta}} - \left(\frac{\theta_r \pi - (1-\gamma)\theta_m \Delta}{2\gamma\theta_r \Delta}\right)$$

$$\underline{p} = \frac{\theta_r^2 - 2\gamma(1-\gamma)(\pi-\Delta)\Delta - \theta_r\sqrt{\theta_r^2 - 4\gamma(1-\gamma)(\pi-\Delta)\pi}}{2(1-\gamma)(\theta_r^2 + \gamma(1-\gamma)\Delta^2)}$$

$$\bar{p} = \min\left\{\frac{\theta_r^2 - 2\gamma(1-\gamma)(\pi-\Delta)\Delta + \theta_r\sqrt{\theta_r^2 - 4\gamma(1-\gamma)(\pi-\Delta)\pi}}{2(1-\gamma)(\theta_r^2 + \gamma(1-\gamma)\Delta^2)}, 1\right\}$$

LEMMA 1. *Suppose that an inclusive club is formed in every period from some finite time T onward. Then an inclusive club forms in every period $t > T$. In addition, whenever an inclusive club forms in every period, $\{p^t\}_{t=T}^{\infty}$ is a strictly decreasing sequence with* $\lim_{t\to\infty} p^t = 0$.

If an inclusive club forms from any time T onward then there is moderation: $\lim_{t\to\infty} p^t = 0$. The reason is that club-level transmission within an inclusive club is the same as group-level transmission, i.e., based on uniform random matching within the identity group. If these were the only two forms of transmission, i.e., the strength of institutional transmission γ equals zero, then cultural transmission would be 'neutral' and the initial distribution of traits would be preserved: $p^t = p^0$ for all t. By assumption, however, $\gamma > 0$. Hence, institutional transmission tilts the cultural transmission process in favour of the moderate trait, driving the radical trait to extinction. As we will now show, radicals need a club composed exclusively of radicals to survive.

LEMMA 2. *Suppose that an exclusive club is formed in every period from some finite time T onward. There exist thresholds $\underline{p}$ and $\bar{p}$ defined in Table 1 such that*

(*i*) *if* $\Delta < \pi - \theta_r^2/[4\pi\gamma(1-\gamma)]$ *then* $\{p^t\}_{t=T}^{\infty}$ *is a strictly decreasing sequence and* $\lim_{t\to\infty} p^t = 0$;
(*ii*) *if* $\Delta \geq \pi - \theta_r^2/[4\pi\gamma(1-\gamma)]$ *then there exist thresholds* $\underline{p} \in (0, 1]$ *and* $\bar{p} \in [\underline{p}, 1]$ *such that*
   (*a*) *if* $p^0 < \underline{p}$ *then* $\{p^t\}_{t=T}^{\infty}$ *is a strictly decreasing sequence and* $\lim_{t\to\infty} p^t = 0$;
   (*b*) *if* $p^0 = \underline{p}$ *and* $T = 0$ *then* $p^t = \underline{p}$ *for all t;*
   (*c*) *if* $p^0 > \underline{p}$ *then the dynamic exhibits radicalism with* $\lim_{t\to\infty} p^t = \bar{p}$.

Therefore, an exclusive club must repeatedly form for radicals to survive. This *cultural purification* of the club restricts club-level transmission to the radical trait, acting as a counterbalance to institutional transmission of the moderate trait.

2.3. *Unintended Radicalisation*

We begin with the benchmark case of a myopic or fully impatient organisation ($\mu = 0$). In this case, the organisation has no forward-looking concerns; it simply sets s^t to maximise total participation X^t in that period (ignoring the effect on p^{t+1}). Hence, if radicalisation occurs, it is *unintentional*. In the next subsection, we set $\mu > 0$ and identify an *intentional* radicalisation strategy. The analysis here references a number of lemmas that are stated and proved in the Appendix.

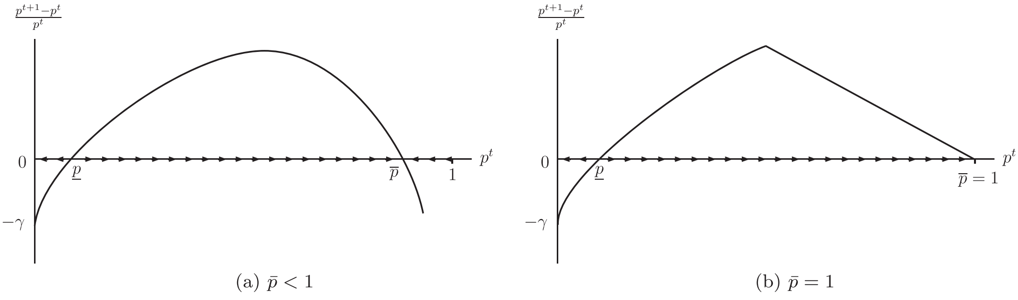

Fig. 2. *The Evolution of the Share of Radical Types p^t under an Exclusive Club for $\Delta \geq \underline{\Delta}$.*
Notes: Panel (a) illustrates the interior case and Panel (b) illustrates the boundary case.

2.3.1. *When does an exclusive club form?*

When the organisation has no forward-looking concerns, it screens out moderates and forms an exclusive club whenever the share of radicals is sufficiently large. Otherwise, it forms an inclusive club.

PROPOSITION 1. *Consider the myopic case $\mu = 0$. There exists a unique IE. In this IE, there exists a threshold share of radicals $\hat{p} \in (0, 1)$ given in Table 1 such that an exclusive club forms in state p if and only if $p > \hat{p}$.*

The proof of this and all other results are in the Appendix.

Moderates value the club good as well as the club's protection from institutional transmission less than radicals. To induce moderates to join, strictness must be set relatively low, $s \leq \underline{s}(p)$. Alternatively, the organisation could raise strictness and elicit larger contributions from the mass p of radicals. An inclusive club's total participation is $\underline{s}(p)$, while an exclusive club's equilibrium output is $p\bar{s}(p)$. Hence, an exclusive club forms when the proportion of radicals is greater than the relative strictness of an inclusive club: $p > \underline{s}(p)/\bar{s}(p)$. This is the case if and only if $p > \hat{p}$. Thus, a growing share of radicals leads to the break up of an inclusive club and the formation of an exclusive club.

2.3.2. *When does an exclusive club produce radicalism?*

Suppose that an exclusive club is formed in every period. Then the dynamics depend crucially on the degree of cultural intolerance Δ. We find that an exclusive club can only produce radicalism if $\Delta > \pi - \theta_r^2/[4\pi\gamma(1-\gamma)]$. In this case, we depict the dynamics in Figure 2 with the thresholds $\underline{p}$ and $\bar{p}$ defined in Table 1 (Lemma 2). There is path dependence. If the initial share of radicals is sufficiently high, i.e., $p^0 > \underline{p}$, the dynamic converges to $\bar{p} \in (0, 1]$ and there is radicalism. If $p^0 < \underline{p}$, however, the share of radicals converges to zero and there is moderation.

Note that, for $\bar{p} < 1$, an exclusive club generates a non-decreasing share of radicals only when p is intermediate, $p \in [\underline{p}, \bar{p}]$. This is for the same reason that $\bar{s}(p)$ can be a non-monotonic function of p, i.e., the countervailing effects of increasing returns to club goods production and cultural substitutability.

2.3.3. *Equilibrium dynamics*

Taken together, these results imply that radicalism depends on (*1*) the formation of an exclusive club, which occurs when $p^0 > \hat{p}$ by Proposition 1, and (*2*) the ability of an exclusive club to

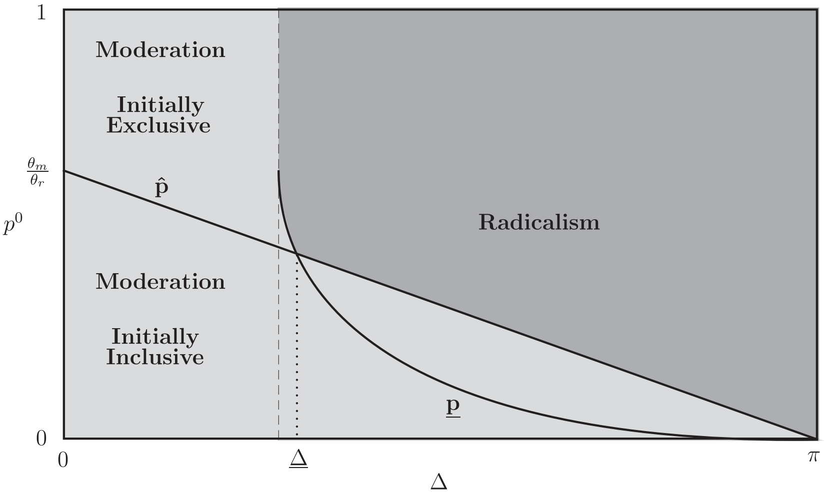

Fig. 3. *Notional Characterisation of the Equilibrium Path in the Myopic Case $\mu = 0$, with $\underline{\Delta} > 0$ and θ_m Close to θ_r.*

stabilise the share of radicals above $\hat{p}$. By inspection of Figure 2, this requires $\hat{p} \in [\underline{p}, \bar{p})$. We can show that $\hat{p} < \bar{p}$ for sufficiently large Δ. This leads to the following result.

PROPOSITION 2. *In the myopic case $\mu = 0$, there exist thresholds $\bar{\Delta} \in [0, \pi)$, $\underline{p} \in (0, 1)$ and $\hat{p} \in (0, 1)$ such that there is radicalism on the equilibrium path if and only if the degree of cultural intolerance is sufficiently large, $\Delta > \bar{\Delta}$, and the initial share of radicals is sufficiently large, $p^0 > \hat{p}$ and $p^0 \geq \underline{p}$.*

The threshold $\bar{\Delta}$ is implicitly defined in the proof. The set of conditions under which there is radicalism, that is,

$$\Omega = \{(\Delta, p^0) \colon \Delta \geq \bar{\Delta}, p^0 > \hat{p} \text{ and } p^0 \geq \underline{p}\}$$

has positive Lebesgue measure because $\bar{\Delta} \in [0, \pi)$, $\underline{p} \in (0, 1)$ and $\hat{p} \in (0, 1)$. Under all other conditions $(\Delta, p^0) \notin \Omega$, the identity group exhibits moderation.

Figure 3 depicts the equilibrium path and its dependence on the conditions (p^0, Δ), i.e., the initial share of radicals and the degree of cultural intolerance. Additional results required to construct the figure are contained in the Appendix (Lemmas 6 to 8). Again, radicalism requires that an exclusive club forms and stabilises the share of radicals above $\hat{p}$. First, $p^0 > \hat{p}$ is a necessary and sufficient condition for an exclusive club to form (Proposition 1). To the contrary, suppose that$p^0 \leq \hat{p}$. Then an inclusive club forms initially and the share of radicals declines. Hence, $p^1 < \hat{p}$, so an inclusive club forms again in period 1. Iterating, an inclusive club forms in every period and moderation occurs: $\lim_{t \to \infty} p^t = 0$. Second, $\Delta > \bar{\Delta}$ and $p^0 \geq \underline{p}$ is a necessary condition for an exclusive club to stabilise p^t above $\hat{p}$ (Lemma 2). Suppose instead that $p^0 \in (\hat{p}, \underline{p})$. An exclusive club forms initially (because $p^0 > \hat{p}$), but the share of radicals declines monotonically (because $p^0 < \underline{p}$) until an inclusive club forms, and moderation occurs.

The final condition for radicalism is $\hat{p} < \bar{p}$, so that p^t is stabilised above $\hat{p}$. This condition is satisfied if and only if Δ is sufficiently large.[12]

Therefore, when Δ is sufficiently large, there exist multiple steady states with radicalism occurring where the initial share of radicals is large enough. In this case, history matters and (otherwise) identical identity groups on either side of the threshold for p^0 diverge, with radical types surviving in one group and becoming extinct in the other.

2.4. *Dynamic Radicalisation Strategy*

The radicalism described so far is unintentional. We are now interested in the dynamic strategies that can be used by a forward-looking organisation ($\mu > 0$). Under certain conditions, the organisation produces radicalisation where it would not in the myopic case. Under other conditions, the organisation produces moderation where it would not in the myopic case. The following proposition relies on two thresholds:

$$\bar{\theta} \equiv \theta_r \min\left\{\frac{\sqrt{p^0}(\pi + \gamma p^0 \Delta)}{\pi - \Delta + (1-\gamma)p^0\Delta}, \frac{\pi \bar{p}}{\pi - \Delta + (1-\gamma)\bar{p}\Delta}\right\}$$

$$\text{and} \quad \underline{\Delta} \in \left[\max\left\{0, \pi - \frac{\theta_r^2}{4\gamma(1-\gamma)\pi}\right\}, \pi\right).$$

PROPOSITION 3. *Consider the forward-looking case $\mu \in (0, 1)$. There is a unique IE. In this IE, the limiting share of radicals is the same as in Proposition 2 except in the following cases.*

(*i*) *If $\theta_m < \bar{\theta}$ then, for each pair (p^0, Δ) such that $p^0 \in (\underline{p}, \hat{p})$ and $\Delta > \underline{\Delta}$, radicalisation occurs if the organisation is sufficiently patient (sufficiently large μ).*

(*ii*) *If $\theta_m > \bar{\theta}$, moderation occurs if the organisation is sufficiently patient (sufficiently large μ).*

The first condition essentially means that θ_m is small relative to θ_r, i.e., radicals have a far stronger demand for meaning and belonging than moderates. In this case, the organisation can screen out moderates by setting strictness high enough that only radicals are willing to join and the limiting participation when forming an exclusive club and generating radicalisation, $\bar{p}\bar{s}(\bar{p})$, is greater than the participation under any strategy that does not produce radicalisation. Hence, for a discount factor μ sufficiently close to one, a forward-looking organisation produces radicalisation where a myopic organisation would not. The second condition is the opposite, i.e., moderates and radicals are similar in their demand for meaning and belonging. In this case, it is either impossible to screen out moderates or the limiting participation when forming an inclusive club and generating moderation, $\underline{s}(0)$, is greater than the participation under any strategy that does not produce moderation. This emphasises the point that radicalism does not necessarily maximise total participation. Hence, for a discount factor μ sufficiently close to one, a forward-looking organisation produces moderation even where a myopic organisation would not.

2.4.1. *How do these dynamic strategies work?*

As with radicalism created unintentionally by a myopic organisation, intentional radicalisation is produced by the cultural purification of the club through the screening out of moderates.

[12] For θ_m close to θ_r, the case depicted in Figure 3, $\hat{p} < \bar{p}$ is satisfied whenever $\Delta > \pi - \theta_r^2/[4\pi\gamma(1-\gamma)]$. Otherwise, the threshold $\bar{\Delta}$ might be larger.

It also produces the same limiting share of radicals, $\bar{p}$. There is, however, one distinguishing feature of intentional radicalisation. Along all other equilibrium paths to radicalism, total club participation is no less than total participation in an inclusive club $\underline{s}(p^t)$ at each time t. According to Proposition 3(i), a forward-looking organisation intentionally produces radicalisation by forming an exclusive club at $p^0 \in (\underline{p}, \hat{p})$, where a myopic organisation would form an inclusive club (because $p < \hat{p}$). Thus, total club participation starts out and remains below $\underline{s}(p^t)$ for some time. However, while screening out moderates lowers current participation, it grows the share of radicals over time (because $p^0 > \underline{p}$) by boosting club-level transmission of the radical trait. Eventually, total club participation rises above $\underline{s}(p^t)$ and remains so. If the organisation is sufficiently forward looking, it is willing to make the short-term sacrifice in club participation for the long-term gain. Hence, the exclusive club can initially appear weaker than an inclusive club, but it grows larger and stricter over time and thus stabilises radicalism. There is an open set of conditions under which this dynamic radicalisation strategy is used, as depicted by the lens-shaped region in Figure 4(a).

Intentional moderation occurs when an inclusive club would currently produce lower participation than an exclusive club. Forming an inclusive club, however, reduces the share of radicals over time. This makes it more valuable for moderates to join the inclusive club, because they are less likely to be matched with radicals within the club during cultural transmission. This in turn enables the organisation to raise the strictness of the inclusive club over time to the point at which it generates greater participation than an exclusive club. There is an open set of conditions under which this dynamic moderation strategy is used, as depicted by the upper-right region in Figure 4(b).

Both the dynamic radicalisation and moderation strategies operate through a form of cultural purification. In the moderation case, the cultural purification is not sudden. Instead, the formation of an inclusive club leads to the gradual elimination of radicals from the group through cultural transmission, culminating in a club composed exclusively of moderates. In the radicalisation case, cultural purification occurs suddenly with the screening out of moderates from the club.

2.5. *Total Radicalisation*

There is a simple additional condition under which radicalisation is total, i.e., $\bar{p} = 1$. It is necessary and sufficient that the strictness of an exclusive club when the entire identity group is composed of radicals is one, i.e., $\bar{s}(1) = 1$. Such an exclusive club is totally isolated from mainstream society. This leads to the following result.

PROPOSITION 4. *Total radicalisation,* $\lim_{t\to\infty} p^t = 1$*, occurs if and only if, in addition to the conditions for radicalisation in Proposition* 2 *or Proposition* 3, $\theta_r \geq \pi - \gamma\Delta$.

Therefore, when radical types place a sufficiently high value θ_r on the club good and radicalisation occurs, moderates are totally eliminated from the identity group. The threshold value is increasing in the returns from outside activity π and decreasing in the ‘expected cost’ to radicals of exposure to institutional transmission $\gamma \times \Delta$.

3. Why do Counter-Radicalisation Policies Fail?

The baseline radicalisation mechanism we have identified—cultural purification—has important implications for counter-radicalisation policies. In this section, we show that a number of conventional policies can backfire and fuel this process of radicalisation. We also consider packages

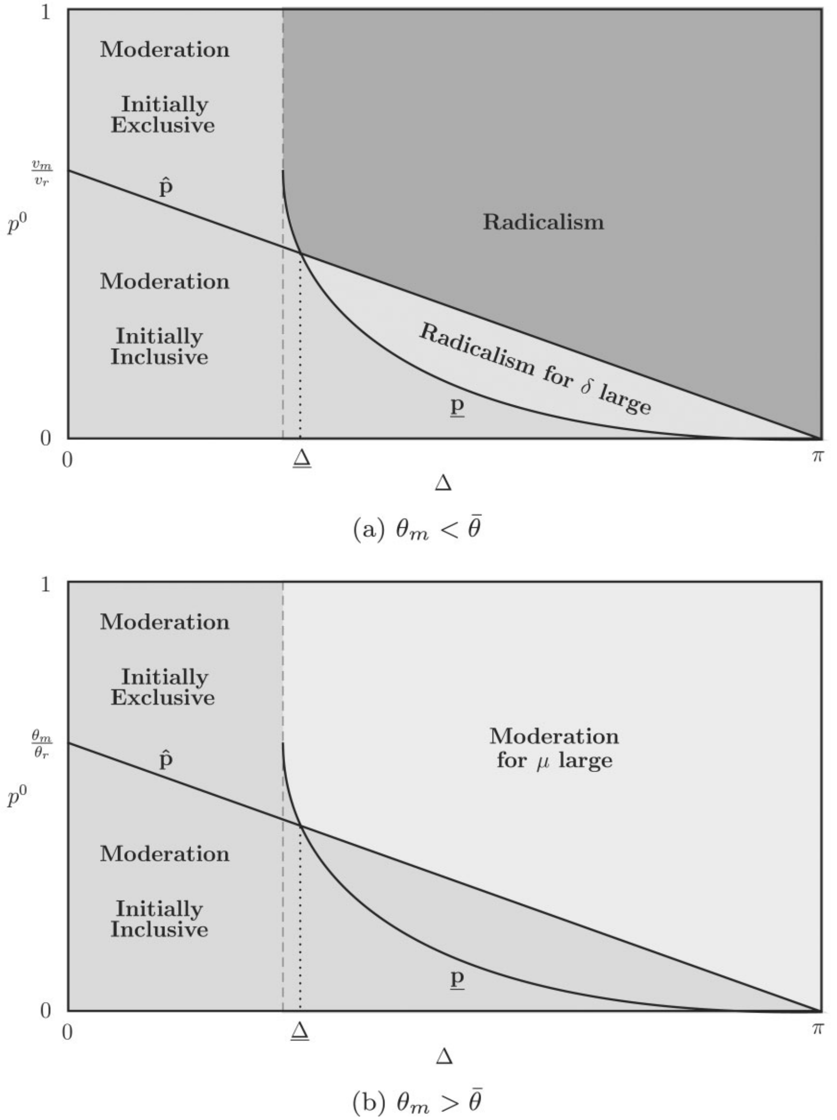

Fig. 4. *Notional Characterisation of the Equilibrium Path in the Forward-Looking Case $\mu > 0$, with $\underline{\Delta} > 0$.*
Notes: Panel (a) illustrates the case in which radicalism can occur and Panel (b) illustrates the case in which it cannot.

of multiple policies. After two additional radicalisation mechanisms are introduced in Section 4, we set out three general principles for counter-radicalisation policy in Section 5.

3.1. *Stigmatising Participation*

The most obvious way to limit the organisation's ability to radicalise is to 'tax' club membership and thereby reduce the value of its platform. This can be done, for example, by stigmatising participation in identity-based activities. The problem with such normative interventions is that the policymaker may not be able to precisely control the degree of stigma. We find that if stigma is not correctly tuned, it can expand the conditions under which radicalisation occurs.

Let c be the fixed cost paid by each club member, which we interpret as stigma. Define

$$\underline{c}(p) = \frac{\theta_m^2}{4(\pi + \gamma p \Delta)} \quad \text{and} \quad \bar{c}(p) = \frac{\theta_r^2 p}{4(\pi - \Delta + (1-\gamma) p \Delta)},$$

where $\underline{c}(p)$ is the maximum cost under which an inclusive club can form and $\bar{c}(p)$ is the maximum cost under which an exclusive club can form. Beyond these costs, all individuals choose non-affiliation. We can now state the following proposition.

PROPOSITION 5. *For all $c > 0$, there is a unique IE. Stigma changes the equilibrium path in the following ways.*

(i) *Suppose that $\underline{c}(p^0) < c \leq \bar{c}(p^0)$.*
  (a) *If $p^0 \in [\underline{p}, \bar{p})$ then radicalisation occurs for all $\mu \in [0, 1)$.*
  (b) *If $p^0 \geq \bar{p}$ and $\underline{c}(\bar{p}) < c \leq \bar{c}(\bar{p})$ then radicalism occurs for all $\mu \in [0, 1)$.*
(ii) *If $c > \bar{c}(p^0)$ then moderation occurs for all $\mu \in [0, 1)$.*

According to part (i) of the proposition, when stigma c lies in an intermediate range, there are conditions under which radicalism occurs where it otherwise would not. Hence, stigmatising club membership backfires. Note that the interval in part (i) is non-degenerate if p^0 is sufficiently large. Specifically, there exists an open set of conditions under which $\underline{c}(p^0) < \bar{c}(p^0)$ with $p^0 < \hat{p}$ (Lemma 9). On the other hand, part (ii) states that if stigma is sufficiently large then moderation is guaranteed, and occurs where there would otherwise be radicalism. Note that, when stigma is low, there is little if any change to the earlier results.

The reason that stigma can backfire is as follows. Consider, for example, the case $p^0 \in [\underline{p}, \hat{p})$. For $c = 0$, we know from Proposition 2 (and by continuity) that a sufficiently impatient/myopic organisation ($\mu \approx 0$) forms an inclusive club in every period and the share of radicals converges monotonically to zero. In contrast, when $\underline{c}(p^0) < c \leq \bar{c}(p^0)$, the degree of stigma makes it impossible to form an inclusive club at $t = 0$. As moderates exit to avoid stigma, the organisation has no choice but to form a strict exclusive club, which can radicalise the identity group (if $\Delta > \bar{\Delta}$), even though it would prefer to form an inclusive club and produce moderation. This applies for all discount factors $\mu \in [0, 1)$. Thus, raising stigma can create a *synthetic radicalisation strategy*.

If stigma were raised very high, however, specifically if $c > \bar{c}(p^0)$, then even radicals would exit the club and moderation would prevail. Thus, stigma is only effective at limiting radicalisation when the policymaker can raise stigma above the level that can be tolerated by radicals. If it cannot guarantee such a high degree of stigma then it may be better off avoiding the stigmatisation of identity-based activities altogether. Our results suggest that an ‘all or nothing’ approach is required for stigma to inhibit, rather than fuel, radicalisation.

The degree to which stigma can fuel radicalisation is illustrated with a numerical example depicted by Figure 5. Under the parameter values stated in the caption, including a discount factor of $\mu = 0$, an inclusive club is formed in every period and radicals are driven to extinction when there is little to no stigma, $c = 0$ and $c = 0.02$. Neither radicalisation nor radicalism occur. Now consider raising stigma slightly to $c = 0.04$. This prevents the formation of an inclusive club,

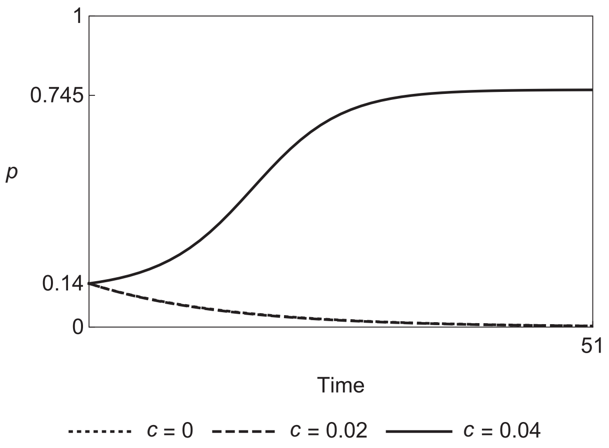

Fig. 5. *Stigma and Radicalisation.*
Notes: Parameters are $\mu = 0$, $\Delta = 1.45$, $\pi = 2$, $\gamma = 0.075$, $\theta_r = 1$, $\theta_m = 0.5$.

but not an exclusive club. Under the new trajectory, the share of radicals increases monotonically over time, along with total participation. In the limit, radicals make up approximately 74.5% of the identity group. Hence, small changes in stigma can lead to large differences in radicalism. That means that the effectiveness of stigmatising policies depends crucially on the policymaker's ability to precisely tune the degree of stigma.

For the remainder of the paper, $c = 0$ except where otherwise stated.

3.2. *Moderating the Moderates*

A second counter-radicalisation strategy is for the policymaker to reduce the demand for meaning and belonging, θ_τ, among members of the identity group. For example, 'hearts and minds' interventions (e.g., community service) are meant to reduce oppositional attitudes among the target group (Kepel, 2004; Berman *et al.*, 2011). While it might appear that no harm could be done, our model suggests that there is a danger with such interventions if they mainly influence moderates, i.e., lower θ_m. Indeed, a policymaker may concede that radicals are likely to be resistant to 'hearts and minds' strategies and focus instead on moderating the moderates in the identity group. Moderates may be viewed as 'the sea in which radicals swim', to paraphrase Mao Tse-Tung (see Berman *et al.*, 2011). According to our model, this can fuel radicalisation.

PROPOSITION 6. *For θ_m sufficiently small (but positive) and $\Delta > \underline{\Delta}$, radicalism occurs for all $p^0 \geq \underline{p}$ and $\mu \in [0, 1)$.*

The reason is as follows. The strictness of an inclusive club is determined by the moderate type's participation constraint. Lowering θ_m lowers the club's value to moderates and thereby lowers the strictness the organisation can demand when forming an inclusive club. If θ_m is reduced too far, the organisation can achieve a higher total payoff by screening out moderates and catering exclusively to radicals at high strictness, regardless of the state p^t. In this case, the only thing stopping radicalism is the inability of an exclusive club to generate radicalism (i.e.,

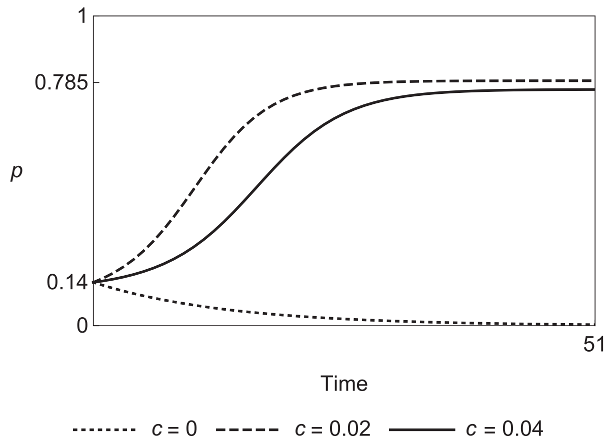

Fig. 6. *Stigma and Radicalisation with θ_m Reduced to* 0.425.
Notes: Parameters are $\mu = 0$, $\Delta = 1.45$, $\pi = 2$, $\gamma = 0.075$, $\theta_r = 1$.

$\Delta < \underline{\Delta}$ or $p^0 < \underline{p}$). This result is stronger than simply pushing θ_m below $\bar{\theta}$ in Proposition 3, because it applies for all discount factors $\mu \in [0, 1)$.[13]

Consider, for example, the case of a religious identity group. According to our results, if the moderates become atheists ($\theta_m = 0$), they will be replaced over time by radicals. In contrast, if they maintain their religiosity, the identity group remains unified and moderate. The reason this insight is surprising is because the crucial role of cultural purification and endogenous club formation in the radicalisation process has not, until now, been well understood.

In addition, the policy of targeting moderates interacts with stigma ($c > 0$) and expands the set of conditions under which stigmatising club membership can backfire. Suppose that $c < \underline{c}(p^0) < \bar{c}(p^0)$, so the organisation can choose between an inclusive club and an exclusive club at $t = 0$. Note that $\underline{c}(p)$ is strictly increasing in θ_m and $\bar{c}(p)$ is independent of θ_m. Hence, lowering θ_m can lead to $\underline{c}(p^0) < c < \bar{c}(p^0)$, ruling out the formation of an inclusive club at $t = 0$ and forcing the organisation to form an exclusive club. Thus, radicalisation can occur even when the organisation prefers inclusivity and moderation. That is, moderating the moderates when there is stigma can activate the synthetic radicalisation strategy. Let us illustrate by returning to the example in Figure 5, except with θ_m lowered to 0.425 (a 15% decrease). The new dynamics are depicted by Figure 6. In particular, we show that the critical intermediate range of stigma expands, with even the lower degree of stigma ($c = 0.02$) producing radicalism, specifically a limiting share of radicals of approximately 78.4%.

3.3. *Intolerance toward Radicals*

A third counter-radicalisation strategy is to induce moderates and radicals in the identity group to be more intolerant toward each other. This is captured in our model by a rise in Δ, i.e., the degree to which a person dislikes having a child of a different type. In particular, there have been calls for

[13] Note that reducing θ_m does not necessarily raise total participation. In particular, it reduces total participation if an inclusive club is preserved. If, however, a marginal drop in θ_m breaks up an inclusive club then this spurs radicalisation and increases total participation in the limit above the level of total participation under the original value of θ_m.

moderates to be more intolerant toward radicals (Genicot, 2022). The idea could be that greater intolerance between moderates and radicals in the group could break up coordination among group members, a kind of 'divide and rule' policy. According to our analysis, this is precisely the wrong approach. By splitting up an inclusive club, a rise in Δ can produce radicalisation (both intentional and unintentional) where it otherwise would not occur.

The mechanism is related to the seminal work by anthropologist Mary Douglas (1966) on cultural purity and danger. Groups are forged by the threat of polluting cultural influences, which creates ambiguity and miscoordination. Purity is restored by carefully defining group boundaries in terms of adherence to sacred beliefs and conventions:

> The sacred must always be treated as contagious because relations with it are bound to be expressed by rituals of separation and demarcation and by beliefs in the danger of crossing forbidden boundaries.
>
> (Douglas, 1966, p. 27)

In our model, the danger posed by 'contagious' cultural traits is captured by the degree of cultural intolerance Δ. The 'separation and demarcation' used to produce cultural purity are equivalent to the formation of an exclusive club. The greater the danger, the larger the impetus for cultural purification of the club through exclusivity. And, as we have seen, the formation of an exclusive club through the screening out of moderates can produce radicalisation.

PROPOSITION 7. *For Δ sufficiently large (but less than π), radicalism occurs for all $p^0 \in (0, 1)$ and $\mu \in [0, 1)$. In addition, the limiting share of radicals $\bar{p}$ is increasing in Δ, strictly so whenever $\bar{p} < 1$.*

We remark that if we were to introduce different levels of intolerance for moderates and radicals, Δ_m and Δ_r, respectively, then the same results hold when raising Δ_m only. That is, making moderates highly intolerant toward radicals is all that is needed to guarantee radicalism.

3.4. *Anti-Radical Messaging*

A fourth counter-radicalisation strategy is to strengthen institutional transmission of the moderate trait, i.e., raise γ. Recall that $\gamma > 0$ means that an inclusive club always produces moderation. Raising γ speeds up this process. On the other hand, it makes club membership more valuable for radicals as a shield from institutional transmission, meaning that the organisation can make stricter demands of an exclusive club. Hence, we find that increasing the strength of institutional transmission up to an intermediate level inhibits radicalisation. But, pushing γ too far can result in a backlash, with a stricter exclusive club forming and fuelling radicalisation.

This is illustrated by Figure 7 that plots the dynamic p^t when an exclusive club is formed. For $\gamma = 0$, it is possible for the organisation to achieve total radicalisation from any $p^0 \in (0, 1)$. A sufficiently patient organisation (μ close to one) would choose to do so. Raising γ to 0.25 shrinks the set of initial conditions from which radicalism can be achieved and rules out total radicalisation. Raising γ further to 0.5 rules out radicalism altogether and guarantees moderation. Hence, choosing an intermediate strength of institutional transmission is an effective counter-radicalisation strategy. However, pushing the strength of institutional transmission further to 0.75 results in backlash, as analysed by Carvalho *et al.* (2022) and studied empirically by Fouka (2020). By increasing the value of an exclusive club to radicals, raising γ makes it again possible to achieve total radicalisation from any initial state p^0 around 0.41 or larger.

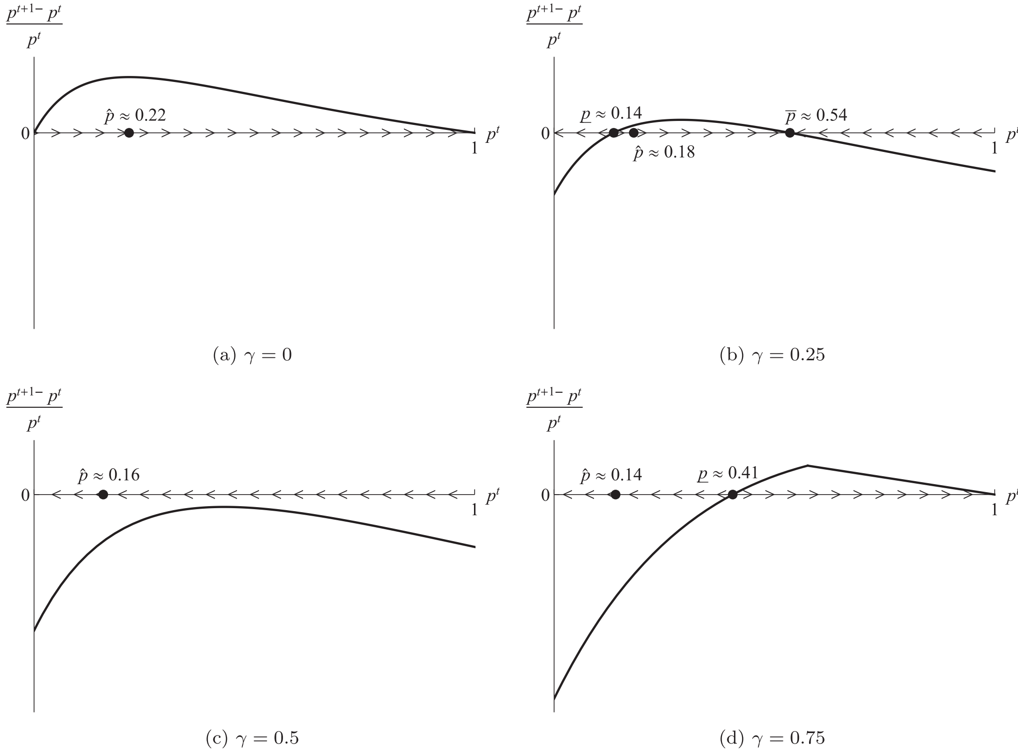

Fig. 7. *Dynamics Under Exclusivity with (a) $\gamma = 0$, (b) $\gamma = 0.25$, (c) $\gamma = 0.5$ and (d) $\gamma = 0.75$.*
Notes: Parameters are $\Delta = 1.45$, $\pi = 2$, $\theta_r = 1$.

We can formalise the effectiveness of an intermediate γ in inhibiting radicalisation, and the danger of pushing institutional transmission too far, by the following proposition.

PROPOSITION 8. *There exist thresholds $\underline{\gamma} \in (0, \frac{1}{2})$ and $\bar{\gamma} \in (\frac{1}{2}, 1)$ such that moderation is guaranteed if $\gamma \in (\underline{\gamma}, \bar{\gamma})$ and $\theta_r < \sqrt{\pi(\pi - \Delta)}$. Furthermore, $\partial\underline{\gamma}/\partial\pi < 0$ and $\partial\bar{\gamma}/\partial\pi > 0$.*
Otherwise, there exist conditions $(\mu, \Delta, \pi, \theta_r, \theta_m)$ under which radicalisation occurs, as described in Propositions 2 *and* 3.

Hence, when radical types are not too radical (θ_r low) and γ is tuned properly, the strength of institutional transmission can rule out radicalism. We remark that to influence the identity group to this degree, the institution doing the transmitting (e.g., mainstream media, public education) must have sufficient legitimacy among group members. It must also have sufficient control of γ to avoid a backlash. These are issues we take up in Section 5.

Before moving on, we can also think about the effect of policy packages. Consider, for example, combining efforts to tune γ with policies that influence π (e.g., efforts to reduce labour market discrimination). Proposition 8 shows that $\partial\underline{\gamma}/\partial\pi < 0$ and $\partial\bar{\gamma}/\partial\pi > 0$. The implication of this result is two-fold. Efforts undertaken to increase returns to effort for members of the identity group expand the set of γ for which moderation is guaranteed. That is, raising π reduces the risk of backlash. Similarly, if at the initial π, $\theta_r \geq \sqrt{\pi(\pi - \Delta)}$, then increasing π can *create* a channel through which institutional transmission guarantees moderation. On the other hand, policies that unintentionally decrease π have the opposite effect, working against this channel and increasing the risk of backlash.

4. Two Further Radicalisation Mechanisms

This section analyses two additional radicalisation mechanisms that have roots in evolutionary theory.

4.1. *Biased Cultural Transmission*

To this point, group-level cultural transmission has been neutral in that individuals are matched uniformly at random with a group member from their parent's generation. Suppose now that (active) club members act as role models with a disproportionate cultural influence. Specifically, assume that the cultural transmission process is as depicted in Figure 1 except that group-level transmission is biased: with probability α, an individual is matched uniformly at random with a club member and acquires their trait; with probability $1 - \alpha$, an individual is matched uniformly at random with a group I member and acquires their trait. The degree of bias in cultural transmission is $\alpha \in [0, 1]$. In the baseline model, $\alpha = 0$. When transmission is biased, non-members are more likely to be culturally influenced by club members than club members are by non-members. The reasons for $\alpha > 0$ include the following.

Proselytism. An important component of identity-based participation is ideological production and persuasion (McBride, 2015). The Reformation and its Puritan offshoots spread via the production and dissemination of literature by zealots using the printing press (Rubin, 2017). Islamist groups in Egypt in the 1980s used books, journals, recordings of influential preachers on cassette tapes and a network of religiously committed lay volunteers who provided club goods (e.g., healthcare) to persuade and recruit members (Wickham, 2002; Clark *et al.*, 2004; Hirschkind, 2006). The Muslim Brotherhood designated the West as *dar al dawa* (land of preaching). Vidino (2020) described their early activities as follows: 'With few funds but plenty of enthusiasm, they published magazines, organised lectures, and carried out all sorts of activities through which they could spread their ideology' (p. 4). The internet has afforded members of extreme groups even greater cultural influence. ISIS used Twitter and Facebook to locate foreign recruits, followed by Skype video calls to screen and groom them. White supremacist groups in the United States radicalise and recruit through social media and online forums. The bias in cultural transmission reflects the fact that the preacher is more likely to influence through social contact than be influenced. Accordingly, Yusuf al Qaradawi, spiritual leader of the Muslim Brotherhood, urges proselytism in the West in the following terms: '[...] conservatism without isolation, and an openness without melting', which has also been described as a 'balance between cultural impermeability and active sociopolitical interaction' (Vidino, 2020, p. 9).

Prestige bias. The cultural influence of active group members could also derive from greater prestige and visibility in the identity group, a classic form of transmission bias in the cultural evolution literature (Boyd and Richerson, 1988; Henrich and Gil-White, 2001). For example, Chudek *et al.* (2012) showed that children are more likely to copy adult role models to whom bystanders in an experiment pay more attention. In the religious context, acts such as prayer, fasting and community service, as well as religious markers (e.g., veiling), give actively religious group members greater visibility and status. As a consequence, they might exert greater influence in the religious socialisation process. For example, Al-Ajarma (2021) described how Muslims who have completed the pilgrimage to Mecca (Hajj), one of the five pillars of Islam, are considered 'by community members to be trustworthy and truthful' and 'are often invited as witnesses to marriages and to act as judges in cases of dispute as their opinions and ideas are highly respected in the community' (p. 2). Of course, the prestige of active members depends on various institutional factors, including group stereotypes, as we discuss in Section 5.

Biased cultural transmission has two effects. First, it can protect the radicals from complete decay. There may now be two asymptotically stable steady states with a positive share of radicals. Under an exclusive club,

$$\begin{aligned} p^{t+1} = {} & p^t(\bar{s}(p^t) + (1-\bar{s}(p^t))(1-\gamma)(\alpha + (1-\alpha)p^t)) \\ & + (1-p^t)(1-\gamma)(\alpha + (1-\alpha)p^t). \end{aligned} \tag{6}$$

In this context we interpret $\underline{p}$, when interior, as the second-largest fixed point of (6). Second, it supercharges the baseline radicalisation process. Forming an exclusive club shields radicals from transmission of the moderate trait, thus stabilising the share of radicals. This vanguard of radicals then rapidly converts moderates through biased group-level transmission due to their disproportionate cultural influence in the group, expanding the set of conditions under which radicalisation occurs.

PROPOSITION 9. *Suppose that $\Delta > \pi - \theta_r^2/[4\gamma(1-\gamma)\pi]$. Under biased transmission, the following statements hold.*

(*i*) *The thresholds $\underline{p}(\alpha)$ and $\bar{\theta}(\alpha)$ such that radicalism occurs for sufficiently large μ are respectively decreasing (strictly so when interior) and non-decreasing in α.*
(*ii*) *If α is sufficiently large and γ is sufficiently small then $\underline{p}(\alpha) = 0$.*
(*iii*) *The threshold $\bar{p}(\alpha)$ to which the share of radicals converges under radicalism is strictly increasing in α when interior.*

Thus, if radicalism occurs at $\alpha = 0$ then radicalism occurs for all $\alpha > 0$, including from initial conditions $p^0 < \underline{p}(0)$.

When transmission is biased, the threshold for a forward-looking organisation to prefer to radicalise rather than moderate, $\bar{\theta}(\alpha)$, is no less than the corresponding threshold in Proposition 3 when $\alpha = 0$, and strictly greater when the screening condition is non-binding. The greater the bias α, the easier it is to satisfy the conditions for radicalisation. The reason is that the limiting share of radicals produced by radicalisation, $\bar{p}(\alpha)$, is strictly increasing in α whenever interior. In addition, an exclusive club can produce radicalism from any initial share of radicals $p^0 > 0$ for large α and small γ, rather than only from $p^0 \geq \underline{p}(0)$, which is positive. Therefore, biased cultural transmission piggybacks on and bolsters the baseline radicalisation mechanism.

4.2. *Discrimination and Niche Construction*

Suppose that the identity group *I* is a subpopulation (e.g., religious/ethnic minority, fringe political faction) being judged by a larger population (outside of *I*) that can discriminate against members of group *I*. We show that the organisation can boost radicalisation by affecting the level of discrimination faced by group members in the society at large. By discrimination, we mean any action that lowers the payoff to outside activity by members of the identity group I. This includes social discrimination, which negatively affects social interactions with outgroup members, and labour market discrimination, which lowers the expected wage faced by group members. Recall that π is the common component of the payoff to outside activity. Now let $\pi = w - \delta$, where $w \geq 1$ could be the real wage and δ is the level of discrimination against group members.[14] This is a form of *blanket discrimination* as δ affects all members of the identity group I, irrespective of whether they join the club and participate in identity-based activities. In contrast, the stigma analysed in Section 3.1 is directed at active club members only.[15]

We uncover a second path to radicalisation by making discrimination δ^t endogenous and studying its co-evolution with p^t. To isolate this channel, we set the bias in cultural transmission to $\alpha = 0$. Let the initial level of discrimination against group members be $\delta^0 = 0$, as in the baseline model without discrimination. The question is whether the organisation can further radicalise the group by inducing a rise in discrimination δ^t over time. We assume that blanket discrimination is determined by two components: the average group participation and the most extreme participation among group members. Specifically, δ^{t+1} is a weighted average of these two components, with weight $\sigma \in [0, 1]$ placed on the most extreme participation:

$$\delta^{t+1} = (1 - \sigma)X^t + \sigma \max_i X^t(i), \tag{7}$$

where $X^t(i)$ is i's level of participation. We interpret σ as the *degree of stereotyping* of the identity group.

Note that participation in the club is time spent separated from the society at large and in distinctive identity-based activities. The first component of discrimination captures society's response to this, with increased participation in identity-based activities, leading to greater discrimination against identity group members. The second component captures the representativeness heuristic (Kahneman and Tversky, 1972) and stereotyping à la Bordalo *et al.* (2016), whereby group members are judged based on their most distinctive behaviour, irrespective of the frequency of this behaviour in the identity group.

We can now state the following result.

PROPOSITION 10. *Under endogenous discrimination, the following statements hold.*

(*i*) *There exists a threshold $\underline{P} < \underline{p}$ such that radicalism occurs in equilibrium from all initial states $p^0 \geq \underline{P}$ if μ and θ_r are sufficiently large.*
(*ii*) *When radicalism occurs, the share of radicals converges to $\bar{P}$, which is greater than the limiting share of radicals $\bar{p}$ without discrimination whenever $\bar{p} < 1$.*
(*iii*) *Threshold $\underline{P}$ is strictly decreasing in the degree of stereotyping σ and $\bar{P}$ is strictly increasing in σ when interior.*

Thus, if the organisation is sufficiently patient and radicals are sufficiently radical, the organisation will produce radicalisation where it would otherwise not occur, i.e., for $p^0 \in [\underline{P}, \underline{p})$. Rather than a policy backfiring, the organisation can itself fuel radicalisation by inducing blanket discrimination against group members. The endogenous generation of blanket discrimination also makes radicalism more extreme than without discrimination. Finally, the set of initial conditions under which radicalisation occurs grows with the degree of stereotyping, σ. This is illustrated

[14] The original assumption that $\theta_m < \pi = w - \delta$ is replaced by $\theta_m < w - 1$.
[15] This follows the distinction made by Bisin *et al.* (2011b) between conditional and unconditional discrimination.

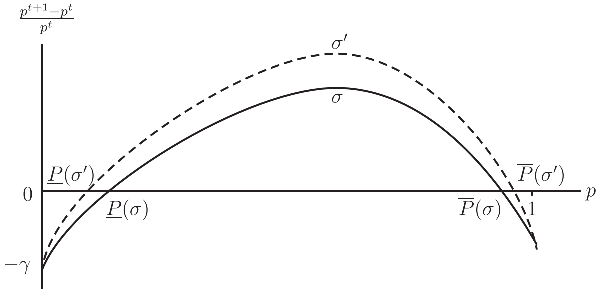

Fig. 8. *The Effect on $\underline{P}$ and $\bar{P}$ of Increasing the Degree of Stereotyping from σ to σ'.*

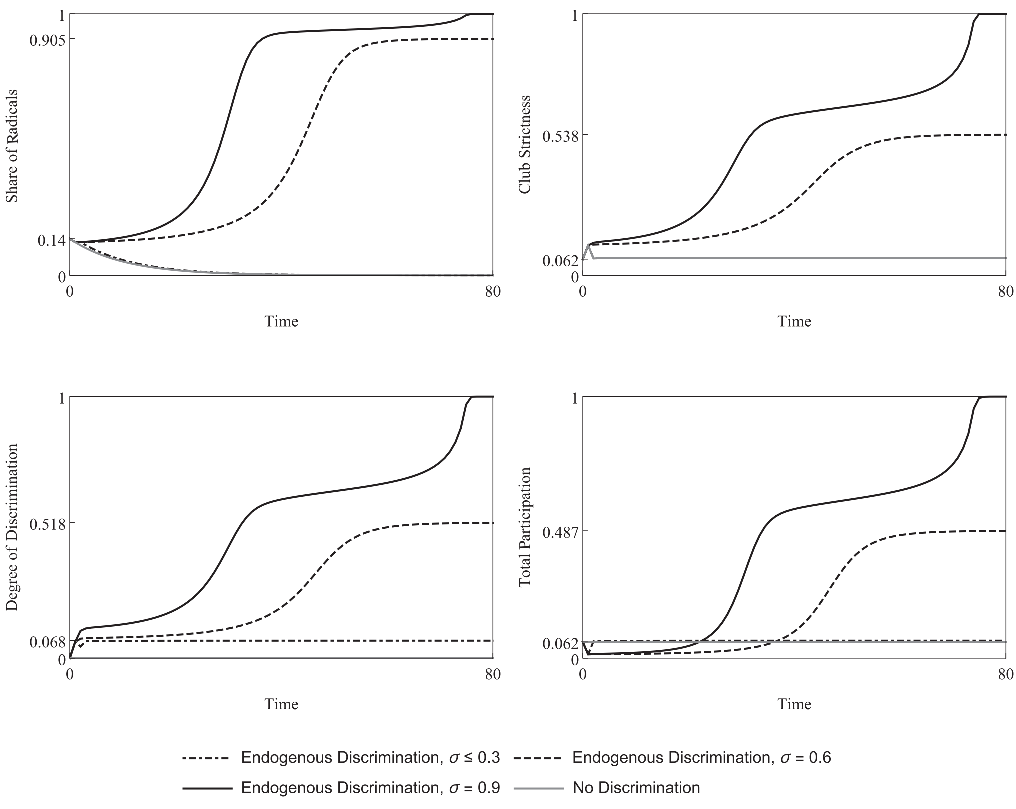

Fig. 9. *Illustration of Radicalisation through Endogenous Discrimination and Niche Construction.*
Notes: Parameters are $w = 2$, $\Delta = 1$, $\gamma = 0.1$, $\theta_r = 1$, $\theta_m = 0.5$, $\delta^0 = 0$ and $\mu \approx 1$.

by Figure 8. An increase in the degree of stereotyping σ lowers the minimum share of radicals necessary for an exclusive club to induce radicalisation, $\underline{P}$, and increases the limiting share of radicals produced by a stable exclusive club, $\bar{P}$.

Numerical examples depicted by Figure 9 illustrate the magnitude of the effects of increasing σ on the share of radicals, club strictness, total participation and discrimination against group members over time. When there is no discrimination, moderation occurs. The same occurs when discrimination is endogenous and the degree of stereotyping, σ, is no greater than 0.3. However, when σ is raised to 0.6, radicalisation occurs. Raising the degree of stereotyping further to 0.9 produces total radicalisation.[16]

The reason discrimination promotes radicalisation is because it lowers the opportunity cost of participation by 'taxing' outside activity for all group members (active or otherwise). This tax boosts future club participation, which further raises blanket discrimination, and so on. We view this as a form of *cultural niche construction* (Odling-Smee *et al.*, 2003). Niche construction is an important factor in evolutionary history. Organisms not only adapt to their environment, but also modify it. In ecology, a classic example is the construction of dams by beavers which alters the operation of natural selection. Laland *et al.* (2001) suggested that cultural niche construction has had an even greater impact on human evolution than gene-based niche construction. The classic example is the effect of domestication of cattle and dairying on the selection of genes for lactose tolerance (e.g., Aoki, 1986). In our model, the organisation shields itself from selection pressures in the form of competition from outside alternatives, by inducing blanket discrimination against group members. This is an economic variant of cultural niche construction. Hauk and Mueller (2015) were perhaps the first to argue that cultural leaders might induce cultural dislike and wage discrimination against group members, what they call a 'cultural alienation' strategy. In their model, discrimination is directly under the control of the cultural leader and fixed for all time. In our dynamic niche construction mechanism, the degree of discrimination δ^t evolves over time with the internal dynamics of the identity group, as well as the degree of stereotyping of the identity group by mainstream society. In particular, there is a novel feedback loop through which the organisation can isolate group members from competing outside opportunities further and further over time.

As the process of radicalisation through niche construction is boosted by a higher degree of stereotyping σ, if societal concerns about growing extremism in the identity group raise σ then they can become self-fulfilling. A process akin to niche construction is described by Adida *et al.* (2016), who presented evidence that a discriminatory equilibrium exists today in France in which Muslims face blanket discrimination for not integrating and do not integrate because they face blanket discrimination. The events of 9/11 and numerous subsequent terrorist attacks raised blanket discrimination against Muslims in Europe and the United States and could have exacerbated this problem. In the United States, 82% of survey respondents believe that there is at least some discrimination against Muslims (Masci, 2019). Hanes and Machin (2014) found that hate crimes in parts of England spiked following 9/11 and the subsequent 7/7 attacks on London, and did not return to pre-attack levels. Our analysis demonstrates how an otherwise temporary spike in blanket discrimination can lock in as part of the process of radicalisation. Accordingly, Connor (2010) showed that Muslim minorities exhibit the highest levels of religiosity in the European countries where they face the greatest discrimination.

5. Policy

We now evaluate existing cultural policies and set out three alternative principles for counter-radicalisation based on our theory. We ground our discussion in the case of social integration

[16] The same patterns emerge under the parametrisation used in Figures 5 and 6 with $p^0 < 0.026 \approx \underline{p}^0$. Decreasing Δ and increasing γ in Figure 9 slows the process down for easier visualisation.

among European Muslims, which is the subject of political attention, including anti-immigration movements and bans on Islamic symbols such as minarets and veils (Wike *et al.*, 2016). Muslims comprise the largest religious minority in most European nations (Pew-Templeton, 2022). Around 6,000 Europeans were estimated to have fought for ISIS in Syria (Kirk, 2016). Within the European Union, they came primarily from Belgium, France, Germany and the UK, the more developed countries. There is evidence that foreign fighters were motivated not by economic deprivation but, by difficulties they had integrating into these societies (Mitts, 2019; Benmelech and Klor, 2020). Thus, it is important to understand the reasons for low socioeconomic integration (see Bisin *et al.*, 2011a; Algan *et al.*, 2012), which in our model is measured by high levels of club participation, s. As we note in the introduction, many other applications of our model could be explored, including contemporary ethnonationalist and populist movements.

5.1. *Approaches to Cultural Diversity*

The literature on approaches to cultural diversity is too vast to survey here.[17] Instead, we focus on what our model has to say about the main approaches.

Multiculturalism. Multicultural policies allow members of ethnic and religious minorities to preserve distinctive beliefs and norms—the so-called 'salad-bowl' approach (e.g., Kukathas, 2003). For example, the Canadian Multiculturalism Act of 1988, which built on the earlier 1971 Multiculturalism Policy, aimed to 'promote the full and equitable participation of individuals and communities of all origins in the continuing evolution and shaping of all aspects of Canadian society and assist them in the elimination of any barrier to that participation' (Government of Canada, 1988). In our model, this means lowering stigma c from participation in identity-based activities. There was good reason for such an approach. Consider, for example, the negative effects of stigma described in interviews by Forstenlechner and Al-Waqfi (2010, p. 767):

> I overheard the owner saying that he has two kinds of Muslims on the payroll: The ones that are integrated and get drunk and eat Bratwurst at company parties and the anti-social ones like myself. So, clearly, in order to be a good Muslim employee, I cannot be a good Muslim (Interviewee 32, Vienna).

We have shown, however, that lowering stigma c can have unintended consequences. By Proposition 5, a high level of stigma c can guarantee moderation and complete social integration: $x(i) = 0$ for all $i \in I$. Lowering stigma to an intermediate level can synthesise a dynamic radicalisation strategy. The organisation would like to form an inclusive club and generate moderation, but moderate types exit to avoid stigma. Thus, the organisation forms an exclusive club and the process of radicalisation through cultural purification is set in motion. If, however, multicultural policies can successfully eliminate stigma then an inclusive and moderate club could be formed.

Assimilationism. In contrast to the multicultural model of inclusion, assimilationist policies aim to create uniformity—the so-called 'melting pot'—by stigmatising and regulating identity-based participation. A prime example is the French secular approach (*laïcité*) that restricts religious expression and identification. France banned conspicuous religious symbols, including the *hijab*, from schools in 2004 and the face veil (*niqab*) from all public spaces in 2011, as did Denmark in 2018. These policies sought to reduce blanket discrimination, while permitting a form of conditional discrimination (stigma) against participants in strict religious groups. A priori, this

[17] For different philosophical approaches to diversity, see, for example, Kymlicka (1995), Appiah (2006), Sen (2006), Muldoon (2016) and Gaus (2019).

might appear to be the right strategy, but a closer look at the dynamics of the model, as well as the reality of terrorism in France, reveals a more complicated picture. As with multicultural policies, *improperly tuned* assimilationist policies backfire and fuel radicalisation. As demonstrated by Proposition 5, raising stigma can make it impossible to form an inclusive club, thereby creating a synthetic radicalisation strategy. Naturally, there are many other factors to consider and one might reasonably support or reject *laïcité* on pre-consequential grounds, for example, as a *sine qua non* of republicanism or as a violation of religious freedom.

Communitarianism. As with multicultural policies, the communitarian approach aims to reduce social and labour market penalties for non-conformity to mainstream norms. It does so through special group-based accommodations, recognising the essential role of communities in shaping preferences and norms (Kymlicka, 1995). An important example is the sharia court system in the United Kingdom that provides Muslims who agree to arbitrate under its conditions with judgements based on sharia law that are fully binding under British law (Lepore, 2012). The obvious risk of such an approach is to lock in members of a religious community and define them primarily in terms of their religious identity. It is thus at odds with liberal individualism. Surprisingly, however, our theory suggests that the communitarian approach could work to reduce radicalisation under certain conditions. If communitarian policies sufficiently raise the value placed on the club good by moderates (θ_m), while holding the value for radicals (θ_r) roughly constant, they disable the dynamic radicalisation strategy described in Proposition 3. If the main effect is to raise the value placed on the club good by radicals (θ_r) then radicalisation is fuelled by communitarian policies.

5.2. *Principles of Counter-Radicalisation*

We now set out three principles for counter-radicalisation policy based on our theory.

(1) Unite and moderate. The first is an overarching principle of Section 3 and, as such, can be stated briefly. If a policymaker cannot eliminate identity-based participation altogether, it must keep moderates and radicals unified in an inclusive club. This lowers strictness and eventually leads to the extinction of radical types. Any policy that breaks up an inclusive club will fuel radicalisation. While it might appear sensible for a policymaker to co-opt moderates in an identity group and isolate radicals, this will force the identity-based organisation to form an exclusive club composed of the remaining radicals. According to our analysis, these radicals might then proceed to convert the moderates and radicalise the identity group. This process is supercharged by biased cultural transmission, which confers greater cultural influence on the radicals, and endogenous discrimination, which adds a further positive feedback in the form of niche construction.

Because cultural purification is the mechanism driving radicalism and radicalisation, a principle of uniting and moderating (alongside the principles discussed below) carries through to environments in which there is competition among clubs rather than a monopoly. Regardless of the number of active clubs, if radicals and moderates are united within a single club, the cultural purification mechanism is disabled. Hence, there is moderation. An inclusive club facing the threat of competition cannot force members to be as strict, further moderating the identity group on the intensive margin with less total participation.[18]

[18] For details on how competition moderates aggregate participation, see Carvalho and Sacks (2021).

(2) The middle path of anti-radical messaging. The state's ability to promote the moderate trait through the media and education system is given by γ, i.e., the likelihood of acquiring the moderate trait through institutional transmission. We have shown that mild anti-radical messaging ($\gamma \approx 1/2$) limits radicalisation. For γ to be greater than zero, the state must have legitimacy among identity group members. Consider the symbiotic relationship between the Egyptian state and the al-Azhar religious establishment (Rubin, 2017). By the mid-1990s, al-Azhar had lost religious legitimacy due its domination by the state since 1952, especially in its support for the Israeli-Palestinian peace process (Oslo Accords). To restore al-Azhar's legitimacy, and its usefulness as a legitimising instrument, the Mubarak government had to allow its clerics to challenge the government, most notably at the 1994 United Nations International Conference on Population and Development held in Cairo (Moustafa, 2000). How European governments might build legitimacy among Muslim minorities, apart from the provision of public goods (e.g., Berman *et al.*, 2011), is difficult to determine. What our theory does say, however, is that extreme anti-radical messaging can backfire by making club membership more valuable to radicals as a shield from institutional transmission. It might be tempting to try to induce rapid assimilation into the host society through extreme cultural interventions. But a backlash may ensue that produces radicalism.

One candidate for an intervention that goes too far is the Danish policy, 'One Denmark without Parallel Societies: No Ghettos in 2030'. In 2018, the Danish government passed a series of measures with the stated goal of more fully integrating non-Western Danish immigrants (Government of Denmark, 2018).[19] Of particular relevance to our analysis are the proposed early-childhood interventions, which strengthen institutional transmission through the state's education system. Starting at age one, all children in *vulnerable residential areas* are required to attend either a public or approved private daycare facility for a minimum of 30 h per week ($\gamma \gg 0$).[20] At these daycare facilities, immigrant children are taught the Danish language and exposed to Danish culture. In each daycare, non-Western immigrants from vulnerable residential areas cannot exceed 30% of their class (Government of Denmark, 2018, p. 8). Parents failing to comply can lose their welfare benefits. Parents taking their children overseas on what is deemed to be a 're-education journey' can face up to four years in prison. Our analysis indicates that such extreme measures can result in a backlash by raising the value of active participation in religious or community activities as a counterbalance to institutional transmission (see also Carvalho *et al.*, 2017; 2022). In our model, radicalisation can ensue, whereas weakening institutional transmission would produce moderation.[21]

(3) Eliminating stereotypes. In the theory of stereotypes introduced by Bordalo *et al.* (2016; 2019), which is based on the representativeness heuristic of Kahneman and Tversky (1972), groups are not judged by the traits/behaviour that are frequently observed in them. Rather they

[19] The document outlining the policy was published only in Danish. Google's machine translation was used to create an English-language version used for this discussion.

[20] Vulnerable residential areas are defined as physically contiguous areas with at least 1,000 residents meeting at least two of the following five criteria: (*i*) the proportion of immigrants and descendants from non-Western countries exceeds 55%, (*ii*) the proportion of residents aged 18–64 who are not actively participating in the labour market or education exceeds 40%, (*iii*) the number of people convicted of violating the Criminal Code, the Weapons Act or the Act on Intoxicating Substances exceeds 2.7%, (*iv*) the proportion of residents aged 30–59 who only have a primary school education exceeds 60% and (*v*) the average gross income between the ages of 15–64 in the area, excluding students, is less than 55% of the average gross income for the same group in the region. A 'ghetto' is defined as a vulnerable residential area that meets at least two of criteria (*i*)–(*iii*) or the share of immigrants and descendants from non-Western countries is over 60% (Government of Denmark, 2018, p. 11).

[21] Other than concerns about liberty, a health impact analysis conducted by Gulis *et al.* (2022) estimates that the policy as a whole will lead to substantial economic losses (equivalent to greater δ). In our model, this also fuels radicalisation.

are judged based on traits/behaviour that are distinctive, i.e., different from other groups, even when they have low frequency within the group. Thus, it could be that a religious minority is judged by the religious participation of its most extreme members. Stereotyping in this way has two effects on radicalisation. Internally, if group members internalise the stereotype, it can raise the cultural influence of radical members within the group ($\alpha > 0$). Externally, stereotyping ($\sigma > 0$) increases blanket discrimination against group members. Hence, an informational intervention that eliminates stereotyping would disable both the biased cultural transmission and niche construction mechanisms.

Consider the dynamic niche construction strategy. Because of stereotyping, society responds to rising participation by radicals in an exclusive club with blanket discrimination against all members, including non-participating moderates. For example, Packer (2015) recounted a discussion on online forums following the Charlie Hebdo attack in Paris: ' "I fear for the Muslims of France. The narrow-minded or frightened are going to dig in their heels and make an *amalgame*" – conflate terrorists with all Muslims'. We have seen how forward-looking identity-based organisations can exploit such biased judgements.[22] Accordingly, Mitts (2019) found an association between local anti-Muslim sentiment and support for ISIS on Twitter by European Muslims. The niche construction strategy is described by former CIA Director General David Petraeus as follows:

> The terrorists' explicit hope has been to try to provoke a clash of civilisations — telling Muslims that the United States is at war with them and their religion. When Western politicians propose blanket discrimination against Islam, they bolster the terrorists' propaganda.
>
> (Petraeus, 2016)

There is no silver-bullet policy for eliminating stereotypes of Muslims. An appropriate bundle of norms and policies might include encouraging greater social contact between Muslims and non-Muslims (Corno *et al.*, 2022), removing stereotypes from portrayals of Muslims in the media and education system, and educational programmes that highlight the secular experiences and achievements of Muslims, and reveal the extent of bias in existing social judgements (Alesina *et al.*, 2018). One concrete policy would be for European governments to abandon their religious co-option strategy in which Islamic leaders are appointed by the state to represent diverse groups of Muslims (including secular individuals). In exchange for market power, these leaders are meant to unite and moderate their communities, by reducing religious strictness s and promoting the government's anti-radical messaging (i.e., raising γ).[23] This approach is reminiscent of co-option strategies used for centuries by rulers in the Muslim world (Auriol and Platteau, 2017; Rubin, 2017). In Europe today, however, the co-option strategy may be counterproductive. According to Sen (2006), after 9/11, the co-option approach acted only to '[...] bolster and strengthen the voice of religious authorities while downgrading the importance of nonreligious institutions and movements' (p. 77). By treating the community as a monolith represented by a religious leader or organisation, the religious co-option strategy could conflate non-practicing and practicing members, thereby increasing stereotyping and fuelling radicalisation.

[22] For the link between terrorism and niche construction, we are grateful for a conversation with Debraj Ray. See also Bueno de Mesquita (2005) and Bueno de Mesquita and Dickson (2007) on how government crackdowns can increase terrorist activity.

[23] Note that co-option may break down when the government loses legitimacy within the religious community, and moderation by community leaders would lead to their loss of legitimacy and influence (e.g., Moustafa, 2000). Hence, at critical junctures, community leaders may be of little use in anti-radical messaging.

Our principle accords with the liberal approach of Sen (2006), which emphasises the multi-dimensional nature of identity and argues against '[...] the odd presumption that the people of the world can be uniquely categorised according to some *singular and overarching* system of partitioning' (p. xii, emphasis in original). He goes on to state: 'The confusion between the plural identities of Muslims and their Islamic identity in particular is not only a deceptive mistake, it has serious implications for policies for peace in the precarious world in which we live' (p. 75).

6. Discussion and Conclusion

By combining a dynamic club model with intergenerational cultural transmission, this paper has identified three mechanisms driving the radicalisation of an identity group. The baseline mechanism relies upon cultural purification of the club, which shields radicals from cultural transmission of the moderate trait and enables them to convert moderates. The baseline process of radicalisation is boosted by two additional mechanisms, bias in cultural transmission and niche construction. We have provided an explanation for why existing counter-radicalisation policies and approaches to cultural diversity can fail, and set out principles for counter-radicalisation policy based on our theory. We have also illustrated the relevance of these results with an application to the social integration and radicalisation of European Muslims.

We now conclude with a discussion of alternative assumptions and future directions for research in this area. The main parametric assumption we make is that a club member's payoff from the club good is proportional to the square root of total club participation. This assumption enabled us to generate clean results and closed-form solutions for the radicalisation thresholds (e.g., $\hat{p}$, $\underline{p}$ and $\bar{p}$). We felt that this was important due to the applied nature of the paper, but radicalisation does not hinge on this functional form. We expect the results to be qualitatively similar, though less precise, for any strictly concave function.

A second important assumption is that the set of types is binary. Again, this is for convenience in generating clean policy-relevant results. With any finite set of types, i.e., θ values, and a suitably modified cultural transmission framework, the same logic would apply and we expect the results to be preserved qualitatively. The same goes for a continuum of types for certain distributions of θ, especially bimodal distributions that approximate the binary set of types assumed in our model. More generally, if there is a sufficiently fat right tail, the organisation will prefer to culturally purify the club by screening out sufficiently moderate types and catering to the more radical types at higher strictness, much as in our binary specification. Then we expect the baseline radicalisation mechanism we have analysed would operate, again in a similar manner to the binary case.

A third assumption is that the bias in cultural transmission α is exogenous. What would a sufficiently forward-looking (high-μ) organisation do if it had total control over α? If the organisation would choose to radicalise for some baseline $\alpha \in (0, 1)$, it would raise α to one and speed up the radicalisation process through faster conversion of moderates by the vanguard of radicals. If the organisation would choose to moderate then it would reduce α to zero to speed up the moderation process. In contrast, if a policymaker had total control over α, it would always set $\alpha = 0$ to inhibit radicalisation or speed up moderation. Future work could examine the endogenous determination of α through, for example, an explicit model of proselytism via social networks.

We have also assumed that the organisation maximises total club participation over time, rather than the size of the club or the welfare of club members. The appropriate objective function is the subject of much debate by economists working on religion. The correct assumption clearly depends on the context. We have assumed that the organisation maximises total participation over time. When it comes to radical movements, our view is that individuals participate because it provides them with a sense of meaning and belonging, as well as other tangible benefits. The organisation, however, may have its own forward-looking objective (e.g., political power) that is furthered by members' participation in identity-based activities. Alternatively, if the organisation's objective were to maximise club size then it would always choose an inclusive club and there would be moderation. If the organisation's objective were to maximise the welfare of club members then the main difference would be that the organisation internalises the cost of participation by members, so strictness would be lower. In addition, the strictness of an inclusive club would be a compromise between the ideal strictness levels of moderates and radicals. Otherwise, the conditions for radicalisation would be qualitatively similar. For example, cultural purification would still occur when the share of radicals is large.

A key feature of our analysis is that it is restricted to radicalisation by a monopolist. Carvalho and Sacks (2021) showed that competition reduces participation in a static religious club model. This bears on a famous debate on religious extremism between David Hume and Adam Smith. Hume proposed that a state-funded religious monopoly would limit religious extremism by bribing the clergy into indolence (Hume, 1983, vol. 3, Section 29). Smith argued to the contrary that religious competition would produce fragmentation and avert struggles for political power among large religious rivals (Smith, 2003).[24] An earlier version of our paper with simplified cultural dynamics showed that competition inhibits radicalisation. With the richer model of cultural transmission found here, the competition case is more complicated and requires a separate paper for a complete analysis. A genuine extension to competition might also examine competing organisations with different objectives. This raises the following questions. What steps can a moderate incumbent take to deter the entry of a radical competitor? Which types of organisations should a policymaker 'target' to inhibit radicalisation of an identity group? Under what conditions could this backfire as targeted organisations convince individuals that this is a blanket attack on the identity group?

In our opinion, the most important direction for future research is empirical work on the radicalisation strategies used by identity-based movements, especially the baseline cultural purification strategy we identify in this paper. Prima facie, many radical movements have followed this trajectory. The Hungarian rabbinate initially welcomed liberalising trends in the nineteenth century, but eventually became the 'cradle of ultra-Orthodoxy' and the most radical anti-assimilationist group in Europe (Carvalho and Koyama, 2016). Italy's Red Brigades expelled moderate members in 1977 in an 'organisational evolution that fostered a violent escalation of armed political activity' (Sundquist, 2010, p. 59). The far-right AfD party in Germany began as a 'softly eurosceptic, non-radical party' that radicalised following the migrant arrivals of 2015 (Arzheimer and Berning, 2019). We expect that our broad predictions will be supported, but that empirical work will also enrich our understanding of the mechanisms identified in this paper. Hopefully, this will lead to another generation of theoretical work, and so forth. Through the necessary dialogue between theoretical and empirical research on radicalisation we might arrive at a deeper understanding of the forces that can destabilise a society and the policies that can be used to avoid identity-based conflict.

[24] See Iannaccone and Berman (2006) for further discussion of this debate.

Appendix A

Proof of Lemma 1. By hypothesis, an inclusive club forms in period T.

By (1), when an inclusive club is formed in period T,

$$\begin{aligned} p^{T+1} &= [s(p^T)p^T + (1 - s(p^T))(1 - \gamma)p^T] \\ &= p^T[1 - \gamma(1 - s(p^T))]. \end{aligned}$$

Hence, $p^{T+1} < p^T$ because $s(p) \leq \underline{s}(p) < 1$ for all p and $\gamma > 0$. Iterating, $\{p^t\}_{t=T}^{\infty}$ is a strictly decreasing sequence. Note that $s(p)$ is maximised at $p = 0$ at $\underline{s}(0) < 1$. Therefore, if an inclusive clubs forms in every period,

$$\lim_{t\to\infty} p^t \leq \lim_{t\to\infty} p^t[1 - \gamma(1 - \underline{s}(0))] = 0.$$

This completes the proof. □

Proof of Lemma 2. By (1), when an exclusive club is formed in period t,

$$p^{t+1} = p^t[\bar{s}(p^t) + (1 - \bar{s}(p^t))(1 - \gamma)p^t] + (1 - p^t)(1 - \gamma)p^t.$$

Hence,

$$\frac{p^{t+1} - p^t}{p^t} = \bar{s}(p^t)[1 - (1 - \gamma)p^t] - \gamma. \tag{A1}$$

Assume for the moment that $\bar{s}(p) < 1$ for all $p \in [0, 1]$. Substituting the expression in (5) for $\bar{s}(p^t) < 1$ into (A1) reveals the right-hand side to be a quadratic function of p^t. If $\Delta < \pi - \theta_r^2/(4\pi\gamma(1 - \gamma))$, there are no real roots. In addition, the right-hand side of (A1) is negative at $p^t = 1$. Therefore, $p^{t+1} < p^t$ for all $p^t \in [0, 1]$. This establishes part (*i*).

If $\Delta > \pi - \theta_r^2/(4\pi\gamma(1 - \gamma))$, there are two real roots, and at least one in [0, 1]. These are $\underline{p}$ and $\bar{p}$ defined in Table 1, with a boundary condition placed on the larger root. If $\Delta = \pi - \theta_r^2/(4\pi\gamma(1 - \gamma))$ then there is one root: $\underline{p} = \bar{p}$.

By (A1), $(p^{t+1} - p^t)/p^t = (1 - \gamma)(1 - p^t) > 0$ when $\bar{s}(p^t) = 1$ and $p^t < 1$, so these roots are unaffected by dropping the assumption that $\bar{s}(p) < 1$ for all $p \in [0, 1]$.

Hence, the dynamics are depicted by Figure 2. For $p^0 > \underline{p}$, the continuity of the right-hand side of (A1) guarantees convergence to $\bar{p}$. □

The proof of Proposition 1 relies on the following three lemmas.

LEMMA 3. *For all* $p \in [0, 1]$, *there exists an* $s > 0$ *such that* $M^*(s, p) = I$.

PROOF. The payoff to moderates from joining an inclusive club exceeds the payoff from non-affiliation if

$$\begin{aligned} \pi(1 - s) + \theta_m(s)^{1/2} &+ [s(1 - p) + (1 - s)(\gamma + (1 - \gamma)(1 - p))]\Delta \\ &\geq \pi + [\gamma + (1 - \gamma)(1 - p)]\Delta. \end{aligned}$$

The equivalent condition for radicals is

$$\pi(1 - s) + \theta_r(s)^{1/2} + [sp + (1 - s)(1 - \gamma)p]\Delta \geq \pi + (1 - \gamma)p\Delta.$$

These conditions simplify to

$$s \leq \left(\frac{\theta_m}{\pi + \gamma p \Delta}\right)^2 \equiv \underline{s}(p) \quad \text{for moderates,} \tag{A2}$$

$$s \leq \left(\frac{\theta_r}{\pi - \gamma p \Delta}\right)^2 \quad \text{for radicals.} \tag{A3}$$

Note that the threshold is positive and lower for moderates because $0 < \theta_m < \theta_r$ by assumption. Hence, for all $p \in [0, 1]$, an inclusive club can be formed at any strictness $s \leq \underline{s}(p)$. □

LEMMA 4. *For each (s, p)-membership game, there is a unique strong Nash equilibrium that respects tie-breaking rule (I).*

PROOF. We have established (in Section 2.1) that $M^*(s, p) \in \{\emptyset, I_r, I\}$. For a given (s, p), there cannot coexist strong equilibria in which $M^*(s, p) = \emptyset$ and $M^*(s, p) = I_r$ that respect tie-breaking rule (I) because radicals either have a profitable coalitional deviation from non-affiliation to joining (in which case $M^*(s, p) \neq \emptyset$), a profitable coalitional deviation from joining to non-affiliation (in which case $M^*(s, p) \neq I_r$) or the tie-breaking rule applies. Likewise, there cannot coexist strong equilibria in which $M^*(s, p) = \emptyset$ and $M^*(s, p) = I$ that respect tie-breaking rule (I). Finally, there cannot coexist strong equilibria in which $M^*(s, p) = I_r$ and $M^*(s, p) = I$ that respect tie-breaking rule (I) because moderates either have a profitable coalitional deviation from non-affiliation to joining (in which case $M^*(s, p) \neq I_r$), a profitable coalitional deviation from joining to non-affiliation (in which case $M^*(s, p) \neq I$) or the tie-breaking rule applies. □

LEMMA 5. *It holds that $p\bar{s}(p)$ is strictly increasing in p.*

PROOF. By (5),

$$p\bar{s}(p) = \begin{cases} p & \text{if } \bar{s}(p) = 1, \\ \left(\dfrac{p\theta_r}{\pi - \Delta + (1-\gamma)p\Delta}\right)^2 & \text{if } 0 < \bar{s}(p) < 1. \end{cases}$$

Thus, $p\bar{s}(p)$ is monotonically increasing whenever $\bar{s}(p) = 1$.

Now suppose that $0 < \bar{s}(p) < 1$. Then,

$$\frac{dp\bar{s}(p)}{dp} = \frac{2\bar{s}(p)(\pi - \Delta)}{\pi - \Delta + (1-\gamma)p\Delta},$$

which is positive because $\pi > \Delta$ by assumption. □

Proof of Proposition 1. By Lemma 4, for each (s, p)-membership game, there is a unique strong Nash equilibrium that respects tie-breaking rule (I). Combined with tie-breaking rule (II) for the organisation, this rules out multiple IE. Hence, the IE we will now characterise is unique.

By hypothesis, the organisation is myopic ($\mu = 0$). Hence, it sets s to maximise $X(s, p)$ in each state p, so $s \in \{\underline{s}(p), \bar{s}(p)\}$. By Lemma 3, for all $p \in [0, 1]$, there exists an $s > 0$ such that $M^*(s, p) = I$. Hence, the organisation will never set s such that $M^*(s, p) = \emptyset$, as $X^*(s, p) = 0$ would reduce its payoff. Thus, the organisation chooses between an inclusive and exclusive club.

By (4) and (5), the organisation optimally forms an inclusive club at strictness $\underline{s}(p)$ and an exclusive club at strictness $\bar{s}(p)$. Note that $X(\underline{s}(p), p) = \underline{s}(p)$ and $X(\bar{s}(p), p) = p\bar{s}(p)$. Therefore, the organisation chooses an inclusive club if

$$\underline{s}(p) \geq p\bar{s}(p). \tag{A4}$$

We claim there exists a threshold $\hat{p} \in (0, 1)$ such that (A4) holds if and only if $p \leq \hat{p}$. This would establish the proposition.

To prove the claim, first note that $\underline{s}(0) > \bar{s}(0)$ by (4) and (5). Second, $\underline{s}(p) < \bar{s}(p) = p\bar{s}(p)$ for $p = 1$ by (4) and (5). Third, $p\bar{s}(p)$ is continuous and strictly increasing in p by Lemma 5. Fourth, by (4), $\underline{s}(p)$ is continuous and strictly decreasing in p. Taken together, these facts imply that there exists a unique threshold $\hat{p} \in (0, 1)$ as described.

Substituting (4) and (5) into $\hat{p} = \underline{s}(\hat{p})/\bar{s}(\hat{p})$ and solving for $\hat{p}$ yields the threshold given in Table 1. □

The proof of Proposition 2 relies on Lemma 2 and the following three lemmas, the first of which is an augmented version of Lemma 1.

LEMMA 1′. *Let $\mu = 0$. Suppose that an inclusive club forms in period T. Then, an inclusive club forms in every period $t > T$. In addition, whenever an inclusive club forms in every period,* $\{p^t\}_{t=T}^{\infty}$ *is a strictly decreasing sequence with* $\lim_{t\to\infty} p^t = 0$.

PROOF. By hypothesis, an inclusive club forms in period T. Hence, strictness $s(p) \leq \underline{s}(p)$ and, by Proposition 1, $p^T \leq \hat{p}$. By Lemma 1, $p^{T+1} < p^T$ because $s(p) \leq \underline{s}(p) < 1$ for all p and $\gamma > 0$. This implies that $p^{T+1} < \hat{p}$, so an inclusive club forms again in period $T + 1$. Iterating, an inclusive club forms in every period $t > T$. The remainder of the proof follows immediately from Lemma 1. □

LEMMA 6. *It holds that $\hat{p}$ is strictly decreasing in Δ on the domain $\Delta \in (0, \pi)$.*

PROOF. By definition, $\hat{p}\bar{s}(\hat{p}) = \underline{s}(\hat{p})$.

Suppose that $\bar{s}(\hat{p}) = 1$. Then,

$$\frac{d\hat{p}}{d\Delta} = \frac{\partial \underline{s}(\hat{p})}{\partial \hat{p}}\frac{d\hat{p}}{d\Delta} + \frac{\partial \underline{s}(\hat{p})}{\partial \Delta}$$

$$\iff \quad \frac{d\hat{p}}{d\Delta}\left(1 - \frac{\partial \underline{s}(\hat{p})}{\partial \hat{p}}\right) = \frac{\partial \underline{s}(\hat{p})}{\partial \Delta} = -\frac{2\gamma\hat{p}\underline{s}(\hat{p})}{\pi + \gamma\hat{p}\Delta} < 0.$$

Therefore, $d\hat{p}/d\Delta < 0$ if and only if

$$\frac{\partial \underline{s}(\hat{p})}{\partial \hat{p}} = -\frac{2\gamma\Delta\underline{s}(\hat{p})}{\pi + \gamma\hat{p}\Delta} < 1,$$

which is true.

Now suppose that $\bar{s}(\hat{p}) < 1$. Then,

$$\frac{d\hat{p}}{d\Delta}\bar{s}(\hat{p}) + \hat{p}\left(\frac{\partial \bar{s}(\hat{p})}{\partial \Delta} + \frac{\partial \bar{s}(\hat{p})}{\partial \hat{p}}\frac{d\hat{p}}{d\Delta}\right) = \frac{\partial \underline{s}(\hat{p})}{\partial \Delta} + \frac{\partial \underline{s}(\hat{p})}{\partial \hat{p}}\frac{d\hat{p}}{d\Delta}$$

$$\iff \quad \frac{d\hat{p}}{d\Delta}\left(\bar{s}(\hat{p}) + \hat{p}\frac{\partial \bar{s}(\hat{p})}{\partial \hat{p}} - \frac{\partial \underline{s}(\hat{p})}{\partial \hat{p}}\right) = \frac{\partial \underline{s}(\hat{p})}{\partial \Delta} - \hat{p}\frac{\partial \bar{s}(\hat{p})}{\partial \Delta}. \tag{A5}$$

Note that

$$\frac{\partial \bar{s}(p)}{\partial \Delta} = \frac{2(1 - (1 - \gamma)p)\bar{s}(p)}{\pi - \Delta + (1 - \gamma)p\Delta} > 0,$$

so the right-hand side of (A5) is strictly negative. Hence, $d\hat{p}/d\Delta < 0$ if

$$\bar{s}(\hat{p}) + \hat{p}\frac{\partial \bar{s}(\hat{p})}{\partial \hat{p}} > 0,$$

or, equivalently,

$$\frac{2(\pi - \Delta)\bar{s}(\hat{p})}{\pi - \Delta + (1-\gamma)\hat{p}\Delta} > 0,$$

which is true for all $\Delta \in (0, \pi)$. Thus, $\hat{p}$ is strictly decreasing in Δ on the domain $\Delta \in (0, \pi)$. □

LEMMA 7. *It holds that $\underline{p}$ is strictly decreasing in Δ on the domain $\Delta \in (\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$. It holds that $\bar{p}$ is increasing in Δ on the same domain, strictly so whenever $\bar{p} < 1$.*

PROOF. By definition, $\underline{p}$ and $\bar{p}$ are respectively given by the smallest and largest $p \in [0, 1]$ that satisfy

$$\bar{s}(p)(1 - (1-\gamma)p) \geq \gamma. \tag{A6}$$

By Lemma 2 and Table 1, $\underline{p} < 1$ on the domain $\Delta \in (\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$, which implies that (A6) holds with equality at $\underline{p}$. We cannot have $\bar{s}(\underline{p}) = 1$; otherwise,

$$1 - (1-\gamma)\underline{p} = \gamma \quad \Longleftrightarrow \quad \underline{p} = 1,$$

a contradiction. Additionally, at $p = 1$, (A6) is $\bar{s}(1)\gamma \geq \gamma$, which, when combined with the fact that the left-hand side of (A6) is decreasing in p for $p \geq \bar{p}$, implies that at $\bar{p}$ (A6) holds with equality.

Let $\xi \in \{\underline{p}, \bar{p}\}$ denote an arbitrary solution to (A6). For the remainder of the proof (and without loss of generality by the argument in Lemma 2), consider $\bar{s}(\xi) < 1$. Implicitly differentiating (A6) at $p = \xi$ yields

$$\left(\frac{\partial \bar{s}(\xi)}{\partial \Delta} + \frac{\partial \bar{s}(\xi)}{\partial \xi}\frac{d\xi}{d\Delta}\right)(1 - (1-\gamma)\xi) = (1-\gamma)\bar{s}(\xi)\frac{d\xi}{d\Delta}$$

$$\Longleftrightarrow \quad \frac{d\xi}{d\Delta}\left((1-\gamma)\bar{s}(\xi) - (1-(1-\gamma)\xi)\frac{\partial \bar{s}(\xi)}{\partial \xi}\right) = \frac{\partial \bar{s}(\xi)}{\partial \Delta}(1 - (1-\gamma)\xi).$$

As the right-hand side is strictly positive, $d\xi/d\Delta < 0$ if and only if

$$(1-\gamma)\bar{s}(\xi) - (1-(1-\gamma)\xi)\frac{\partial \bar{s}(\xi)}{\partial \xi} < 0. \tag{A7}$$

Substituting

$$\frac{\partial \bar{s}(\xi)}{\partial \xi} = \frac{(\pi - \Delta)\xi^{-1} - (1-\gamma)\Delta}{\pi - \Delta + (1-\gamma)\xi\Delta}\bar{s}(\xi)$$

into (A7) and simplifying yields

$$(1-\gamma)(2\pi - \Delta)\xi - (\pi - \Delta) < 0. \tag{A8}$$

First, consider $\xi = \underline{p}$. As the left-hand side of (A8) is strictly increasing in ξ, it is sufficient to evaluate (A8) at the maximal interior $\underline{p}$. By Table 1, $\underline{p}$ attains this maximal value at $\theta_r^2 = 4\gamma(1-\gamma)\pi(\pi-\Delta)$, which yields

$$\underline{p} = \frac{\pi-\Delta}{(1-\gamma)(2\pi-\Delta)}.$$

Making this substitution in (A8) yields

$$(1-\gamma)(2\pi-\Delta)\frac{\pi-\Delta}{(1-\gamma)(2\pi-\Delta)} = \pi-\Delta.$$

Hence, $\underline{p}$ is strictly decreasing in Δ whenever $\theta_r^2 > 4\gamma(1-\gamma)(\pi-\Delta)\pi$, or, equivalently, when $\Delta \in (\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$. As the maximum value of $\underline{p}$ corresponds to the minimum value of $\bar{p}$, (A7) cannot be satisfied for $\xi = \bar{p}$. Hence, $\bar{p}$ is strictly increasing in Δ whenever $\bar{p} < 1$ and $\Delta \in (\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$. □

Proof of Proposition 2. By Lemmas 1′ and 2, radicalism occurs only if an exclusive club forms in every period. By Proposition 1, this occurs only if $p^t > \hat{p}$ for all t under an exclusive club.

By Lemma 2, $p^t > \hat{p}$ for all t under an exclusive club if and only if $\Delta \geq \pi - \theta_r^2/(4\pi\gamma(1-\gamma))$, $p^0 \geq \underline{p}$, $p^0 > \hat{p}$ and $\bar{p} > \hat{p}$, where $\underline{p}$ and $\bar{p}$ are defined in Table 1.

By Lemmas 6 and 7, $\hat{p}$ is decreasing in Δ while $\bar{p}$ is non-decreasing in Δ. Furthermore, by inspection of Table 1, $\hat{p} = 0$ and $\bar{p} > 0$ at $\Delta = \pi$. Hence, $\bar{p} > \hat{p}$ is satisfied if and only if Δ is above a threshold that is less than π.

Taken together, these facts imply that there exist thresholds $\bar{\Delta} \in [0, \pi)$ and $\underline{p} \in (0, 1)$ such that there is radicalism on the equilibrium path if and only $\Delta \geq \bar{\Delta}$, $p^0 > \hat{p}$ and $p^0 \geq \underline{p}$. □

To guarantee that the interval in Proposition 3(i) is non-degenerate, we prove the following lemma.[25]

LEMMA 8. *There exists a cutoff* $\underline{\Delta} \in [\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$ *such that* $\underline{p} < \hat{p}$ *if* $\Delta > \underline{\Delta}$.

PROOF. Note by Table 1 that $\hat{p} = \underline{p} = 0$ at $\Delta = \pi$ and

$$\lim_{\Delta\to\pi} \frac{d\hat{p}}{d\Delta} < \lim_{\Delta\to\pi} \frac{d\underline{p}}{d\Delta} = 0.$$

Therefore, in the neighbourhood of $\Delta = \pi$, $\underline{p} < \hat{p}$.

In additon, by Lemma 2, $\underline{p}$ is only well defined for $\Delta \geq \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)$.

Taken together, these results imply that there exists a threshold $\underline{\Delta} \in [\max\{0, \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)\}, \pi)$ such that $\underline{p} < \hat{p}$ for $\Delta \in [\underline{\Delta}, \pi)$. □

Proof of Proposition 3. Uniqueness is guaranteed by the same argument used in the proof of Proposition 1.

The conditions other than those in (i) and (ii) are covered by Lemmas 1′, 2(i), 2(ii)(a) and 2(ii)(c). In each case, moderation is unavoidable. Hence, the outcome is unchanged for a forward-looking organisation ($\mu > 0$).

[25] We can also show that the condition in the lemma is necessary. This proof is available on request.

Now consider the exceptional cases in (*i*) and (*ii*). We make use of the following two comparisons:

$$
\begin{aligned}
& \underline{s}(p^0) < \bar{s}(p^0) \\
\iff \quad & \left(\frac{\theta_m}{\pi + \gamma p^0 \Delta}\right)^2 < \left(\frac{\theta_r \sqrt{p^0}}{\pi - \Delta + (1-\gamma) p^0 \Delta}\right)^2 \\
\iff \quad & \theta_m < \theta_r \frac{\sqrt{p^0}(\pi + \gamma p^0 \Delta)}{\pi - \Delta + (1-\gamma) p^0 \Delta}
\end{aligned} \tag{A9}
$$

and

$$
\begin{aligned}
& \underline{s}(0) < \bar{p}\bar{s}(\bar{p}) \\
\iff \quad & \left(\frac{\theta_m}{\pi}\right)^2 < \left(\frac{\bar{p}\theta_r}{\pi - \Delta + (1-\gamma) \bar{p} \Delta}\right)^2 \\
\iff \quad & \theta_m < \theta_r \frac{\bar{p}\pi}{\pi - \Delta + (1-\gamma) \bar{p} \Delta}.
\end{aligned} \tag{A10}
$$

(*i*) By hypothesis, (A9) and (A10) hold, so $\underline{s}(p^0) < \bar{s}(p^0)$ and $\underline{s}(0) < \bar{p}\bar{s}(\bar{p})$. The first condition allows the screening strategy to function. Otherwise, moderates would join the club intended only for radicals and screening cannot occur. The second guarantees that the limiting participation under radicalism exceeds the limiting participation under moderation. In addition, it is hypothesised that $p^0 \in (\underline{p}, \hat{p})$ and $\Delta > \underline{\Delta}$ so that $\underline{p}$ exists by Lemma 2 and $\underline{\Delta}$ exists by Lemma 8.

We proceed by constructing a strategy that produces radicalisation ($p^t \to \bar{p}$). We then show that, for μ close to (but less than) one, the organisation prefers this strategy to any strategy that does not produce radicalisation.

For the radicalising strategy, consider forming an exclusive club in every state $p \in [p^0, \bar{p}]$. Given $p^0 > \underline{p}$ by hypothesis, p^t converges monotonically to $\bar{p}$ by Lemma 2. Denote the induced path from p^0 to $\bar{p}$ by z^t. The organisation's continuation payoff at some finite time T is

$$
\sum_{t=T}^{\infty} \mu^{t-1} z^t \bar{s}(z^t). \tag{A11}
$$

Now consider a strategy that does not produce radicalisation. This means that, for all finite T, an inclusive club must be chosen in at least one period $t \geq T$. Denote the induced path from p^0 by y^t. Since an inclusive club is selected, $z^t \geq y^t$ for all $t \geq T$, and strictly so for some $t \geq T$. This combined with the fact that $p\bar{s}(p)$ is strictly increasing in p and $\underline{s}(p)$ is strictly decreasing in p means that the organisation's continuation payoff at time T is strictly less than

$$
\sum_{t=T}^{\infty} \mu^{t-1} \max\{\underline{s}(0), z^t \bar{s}(z^t)\}.
$$

For T large (but finite), this is no greater than (A11), because $z^t \bar{s}(z^t)$ is approximately $\bar{p}\bar{s}(\bar{p})$, which is greater than $\underline{s}(0)$ by (A10).

For discount factor μ sufficiently large (but less than one), the optimal strategy from $t = 0$ is that yielding the higher continuation payoff from T onward. Therefore, there is radicalisation in equilibrium.

(*ii*) By hypothesis, the inequality in (A10) is reversed, so $\underline{s}(0) > \bar{p}\bar{s}(\bar{p})$.

We proceed in a similar manner, by constructing a strategy that produces moderation ($p^t \to 0$) and showing that the organisation prefers this to any strategy that does not produce moderation for μ close to (but less than) one.

For the moderating strategy, consider forming an inclusive club in every period with the maximal strictness $\underline{s}(p)$. By Lemma 1, p^t converges monotonically to zero (i.e., moderation). Denote the induced path from p^0 to zero by z^t. The organisation's continuation payoff at some finite time T is

$$\sum_{t=T}^{\infty} \mu^{t-1}\underline{s}(z^t). \tag{A12}$$

Now consider a strategy that does not produce moderation. Denote the induced path from p^0 by y^t. The organisation's continuation payoff at time T is no greater than

$$\sum_{t=T}^{\infty} \mu^{t-1}\max\{\underline{s}(y^t),\, y^t\bar{s}(y^t)\}. \tag{A13}$$

By construction, there exists a finite time T such that $y^t > z^t$ for all $t \geq T$. This combined with the fact that $\underline{s}(p)$ is strictly decreasing in p means that $\underline{s}(y^t) < \underline{s}(z^t)$ for all $t \geq T$. In addition, for large t, an upper bound on $y^t\bar{s}(y^t)$ is approximately $\bar{p}\bar{s}(\bar{p})$ by Lemmas 1 and 2. By hypothesis, $\underline{s}(0) > \bar{p}\bar{s}(\bar{p})$. Hence, by the continuity of $\underline{s}(p)$ and $\bar{s}(p)$, for sufficiently large t, $\underline{s}(z^t) > y^t\bar{s}(y^t)$. Taken together, these facts imply that (A12) is greater than (A13).

For discount factor μ sufficiently large (but less than one), the optimal strategy from $t = 0$ is that yielding the higher continuation payoff from T onward. Therefore, there is moderation in equilibrium. □

Proof of Proposition 4. We know that if radicalisation occurs, p^t converges to $\bar{p}$. Hence, we need to show that $\bar{p} = 1$ if and only if $\theta_r \geq \pi - \gamma\Delta$.

Recall that $\bar{p}$ is given by the largest p that satisfies

$$\bar{s}(p)(1 - (1-\gamma)p) \geq \gamma. \tag{A14}$$

The first step is to show that $\bar{p} = 1$ if and only if $\bar{s}(1) = 1$.

If $\bar{s}(1) = 1$ then (A14) holds with equality at $p = 1$. Hence, $\bar{p} = 1$.

If $\bar{s}(1) < 1$ then (A14) evaluated at $p = 1$ is

$$\bar{s}(1)\gamma \geq \gamma,$$

which is violated. Hence, $\bar{p} < 1$ in this case.

The next step is to show that $\bar{s}(1) = 1$ if and only if $\theta_r \geq \pi - \gamma\Delta$. By (5), $\bar{s}(p) = 1$ whenever

$$\left(\frac{\theta_r\sqrt{p}}{\pi - \Delta + (1-\gamma)p\Delta}\right)^2 \geq 1$$

$$\iff \quad \theta_r\sqrt{p} \geq \pi - \Delta + (1-\gamma)p\Delta.$$

At $p = 1$, the above reduces to $\theta_r \geq \pi - \gamma\Delta$, completing the proof. □

To guarantee that the interval $(\underline{c}(p^0), \bar{c}(p^0))$ exists for a non-degenerate subset of the parameter space as described in Proposition 5, we prove the following lemma.

LEMMA 9. *Define*

$$p' \equiv -\frac{\theta_r^2\pi - \theta_m^2(1-\gamma)\Delta}{2\theta_r^2\gamma\Delta} + \sqrt{\left(\frac{\theta_r^2\pi - \theta_m^2(1-\gamma)\Delta}{2\theta_r^2\gamma\Delta}\right)^2 + \frac{\theta_m^2(\pi-\Delta)}{\theta_r^2\gamma\Delta}} \in (0,1).$$

If $\theta_r\sqrt{p'} < (\pi - \Delta + (1-\gamma)p'\Delta)$ *then,* $p' < \hat{p}$ *and* $\underline{c}(p) < \bar{c}(p)$ *for all* $p > p'$.

PROOF. By an analogous argument to Lemma 3, an inclusive club is formed whenever the payoff to moderates from affiliating exceeds their payoff from non-affiliation:

$$\begin{aligned} &\pi(1-s) + \theta_m(s)^{1/2} + [s(1-p) + (1-s)(\gamma + (1-\gamma)(1-p))]\Delta - c \\ &\quad \geq \pi + [\gamma + (1-\gamma)(1-p)]\Delta \\ &\iff \quad -(\pi + \gamma p\Delta)s + \theta_m(s)^{1/2} - c \geq 0. \end{aligned}$$

The left-hand side is maximised at

$$s_m = \frac{1}{4}\left(\frac{\theta_m}{\pi + \gamma p\Delta}\right)^2,$$

yielding a maximum of

$$\frac{\theta_m^2}{4(\pi + \gamma p\Delta)} - c.$$

Similarly, an exclusive club consisting of only radicals can form only if

$$\begin{aligned} &\pi(1-s) + \theta_r(ps)^{1/2} + [s + (1-s)(1-\gamma)p]\Delta - c \geq \pi + (1-\gamma)p\Delta, \\ &\iff \quad -(\pi - \Delta + (1-\gamma)p\Delta)s + \theta_r(ps)^{1/2} - c \geq 0. \end{aligned}$$

The left-hand side is maximised at

$$s_r = \frac{1}{4}\left(\frac{\theta_r\sqrt{p}}{\pi - \Delta + (1-\gamma)p\Delta}\right)^2,$$

yielding a maximum of

$$\frac{\theta_r^2 p}{4(\pi - \Delta + (1-\gamma)p\Delta)} - c.$$

Therefore, an inclusive club can only form if

$$c \leq \frac{\theta_m^2}{4(\pi + \gamma p\Delta)} \equiv \underline{c}(p)$$

and an exclusive club can only form if

$$c \leq \frac{\theta_r^2 p}{4(\pi - \Delta + (1-\gamma)p\Delta)} \equiv \bar{c}(p).$$

Observe that $\underline{c}(p)$ is strictly decreasing in p while $\bar{c}(p)$ is strictly increasing in p. At $p = 0$,

$$\underline{c}(0) = \frac{\theta_m^2}{4\pi} > 0 = \bar{c}(0).$$

At $p = 1$,

$$\underline{c}(1) = \frac{\theta_m^2}{4(\pi + \gamma\Delta)} < \frac{\theta_r^2}{4(\pi - \gamma\Delta)} = \bar{c}(1),$$

so $\underline{c}(p) < \bar{c}(p)$ for all

$$p > -\frac{\theta_r^2\pi - \theta_m^2(1-\gamma)\Delta}{2\theta_r^2\gamma\Delta} + \sqrt{\left(\frac{\theta_r^2\pi - \theta_m^2(1-\gamma)\Delta}{2\theta_r^2\gamma\Delta}\right)^2 + \frac{\theta_m^2(\pi - \Delta)}{\theta_r^2\gamma\Delta}} \equiv p' \in (0, 1),$$

where p' is the value of $p \in (0, 1)$ that solves $\underline{c}(p) = \bar{c}(p)$. We now show that $p' < \hat{p}$. As before, $\hat{p}$ is defined by the p that solves $\underline{s}(p, c) = p\bar{s}(p, c)$, where $\underline{s}(p, c)$ and $\bar{s}(p, c)$ are now respectively defined by

$$\underline{s}(p, c) = \left(\frac{\theta_m}{2(\pi + \gamma p\Delta)} + \sqrt{\left(\frac{\theta_m}{2(\pi + \gamma p\Delta)}\right)^2 - \frac{c}{\pi + \gamma p\Delta}}\right)^2,$$

$$\bar{s}(p, c) = \left(\frac{\theta_r\sqrt{p}}{2(\pi - \Delta + (1-\gamma)p\Delta)} + \sqrt{\left(\frac{\theta_r\sqrt{p}}{2(\pi - \Delta + (1-\gamma)p\Delta)}\right)^2 - \frac{c}{\pi - \Delta + (1-\gamma)p\Delta}}\right)^2,$$

which is less than 1 at $p = p'$ by hypothesis (as $\theta_r\sqrt{p'} < (\pi - \Delta + (1-\gamma)p'\Delta)$). Set $p = p'$, so $\underline{c}(p') = \bar{c}(p')$, which can be rewritten as

$$\left(\frac{\theta_m}{\theta_r}\right)^2 \left(\frac{1}{p'}\right) \frac{1}{\pi + \gamma p'\Delta} = \frac{1}{\pi - \Delta + (1-\gamma)p'\Delta}.$$

Therefore,

$$
\begin{aligned}
p'\bar{s}(p',c) &= \left(\frac{\theta_r p'}{2(\pi-\Delta+(1-\gamma)p'\Delta)}\right. \\
&\qquad \left. + \sqrt{\left(\frac{\theta_r p'}{2(\pi-\Delta+(1-\gamma)p'\Delta)}\right)^2 - \frac{cp'}{\pi-\Delta+(1-\gamma)p'\Delta}}\right)^2 \\
&= \left(\left(\frac{\theta_m}{\theta_r}\right)\frac{\theta_m}{2(\pi+\gamma p'\Delta)}\right. \\
&\qquad \left. + \sqrt{\left(\frac{\theta_m}{\theta_r}\right)^2\left(\frac{\theta_m}{2(\pi+\gamma p'\Delta)}\right)^2 - \left(\frac{\theta_m}{\theta_r}\right)^2\left(\frac{1}{p'}\right)\frac{cp'}{\pi+\gamma p'\Delta}}\right)^2 \\
&= \left(\frac{\theta_m}{\theta_r}\right)^2\left(\frac{\theta_m}{2(\pi+\gamma p'\Delta)} + \sqrt{\left(\frac{\theta_m}{2(\pi+\gamma p'\Delta)}\right)^2 - \frac{c}{\pi+\gamma p'\Delta}}\right)^2 \\
&= \left(\frac{\theta_m}{\theta_r}\right)^2 \underline{s}(p',c) \; < \underline{s}(p',c).
\end{aligned}
$$

As $p\bar{s}(p,c)$ is strictly increasing in p and $\underline{s}(p,c)$ is strictly decreasing in p, $p' < \hat{p}$. □

Proof of Proposition 5. Uniqueness is guaranteed by the same argument used in the proof of Proposition 1.

We consider the following two cases.

(*i*) $\underline{c}(p^0) < c \leq \bar{c}(p^0)$. By construction, the organisation cannot form an inclusive club, but can form an exclusive club. It will choose an exclusive club over inducing non-affiliation of all group members because the latter minimises its payoff. Hence, an exclusive club forms at $t = 0$.

(a) If $p^0 \in [\underline{p}, \bar{p})$ then $p^1 \geq p^0$ by Lemma 2. As $\underline{c}(p)$ is strictly decreasing in p and $\bar{c}(p)$ is strictly increasing in p, $\underline{c}(p^1) < c \leq \bar{c}(p^1)$. Hence, an exclusive club forms again at $t = 1$. Iterating this argument establishes radicalism.

(b) If $p^0 > \bar{p}$ then if an exclusive club were to form in every period, p^t would decline monotonically to $\bar{p}$ by Lemma 2. Hence, if $\underline{c}(\bar{p}) < c \leq \bar{c}(\bar{p})$, the organisation would choose this path over non-affiliation, which minimises its payoff.

(*ii*) $c > \bar{c}(p^0)$. By construction of $\bar{c}(p)$, an exclusive club cannot form at $t = 0$; either group members choose non-affiliation or an inclusive club forms. In both cases, $p^1 < p^0$. As $\bar{c}(p)$ is strictly increasing in p, $c > \bar{c}(p^1)$. Hence, we can iterate this argument to establish that moderation occurs. □

Proof of Proposition 6. Suppose that $\Delta > \pi - \theta_r^2/(4\pi\gamma(1-\gamma))$, so $\underline{p}$ and $\bar{p}$ are well defined. By Table 1, $\underline{p}$ and $\bar{p}$ are unaffected by θ_m. We now show that the organisation will prefer to form an exclusive club in every period when θ_m is sufficiently small. Recall that the maximal per-period payoff from an inclusive club $\underline{s}(p)$ is strictly (and continuously) decreasing in p and the per-period payoff from an exclusive club $p\bar{s}(p)$ is strictly (and continuously) increasing in p. Therefore, the upper bound on the maximal per-period payoff (with respect to state p) from an inclusive club is $\underline{s}(0)$.

Beginning in state $p^0 \geq \underline{p}$, if an exclusive club is formed in every period then, by Lemma 2, $p^t > 0$ for all t. Let z^t denote the time path and $\underline{z} = \inf\{z^t : t \in [0, \infty)\}$ denote the minimal share of radicals along all possible paths z^t such that $p^0 \geq \underline{p}$. Then, the per-period payoff from an exclusive club is bounded below by $p^0\bar{s}(p^0)$. The upper bound of the maximal per-period payoff of an inclusive club is less than the lower bound of the per-period payoff from an exclusive club for all t when

$$\begin{aligned}\underline{s}(0) &< p^0\bar{s}(p^0),\\ \left(\frac{\theta_m}{\pi}\right)^2 &< \left(\frac{\theta_r p^0}{\pi - \Delta + (1-\gamma)p^0\Delta}\right)^2,\\ \theta_m &< \theta_r \frac{\pi p^0}{\pi - \Delta + (1-\gamma)p^0\Delta}.\end{aligned}$$

Therefore, if the above inequality is satisfied (θ_m is sufficiently small) then the organisation chooses to form an exclusive club in every period for all $p^0 \geq \underline{p}$ and $\mu \in [0, 1)$. □

Proof of Proposition 7. Suppose that Δ is sufficiently large (but less than π). By Table 1, $\underline{p}$ and $\bar{p}$ are continuous in Δ. In addition, as $\Delta \to \pi$, $\underline{p} \to 0$ and $\bar{p}$ converges to a positive number. Hence, by Lemma 2, if $p^0 > 0$ and an exclusive club is formed in every period then $p^t \to \bar{p} > 0$ (i.e., radicalism).

We now show that the organisation will prefer to form an exclusive club in every period.

Recall that the per-period payoff from an inclusive club $\underline{s}(p)$ is continuous and strictly decreasing in p and the per-period payoff from an exclusive club $p\bar{s}(p)$ is continuous and strictly increasing in p. Therefore, the maximal per-period payoff (maximised with respect to state p) from an inclusive club is $\underline{s}(0) = (\theta_m/\pi)^2$. Beginning in state p^0, if an exclusive club is formed in every period then $p^t > 0$ for all t. Let z^t denote the time path and $\underline{z} = \inf\{z^t : t \in [0, \infty)\} > 0$ denote the minimal share of radicals along this path. Then the minimal per-period payoff from an exclusive club is $\underline{z}\bar{s}(\underline{z})$ given by (5), which converges to $(\theta_r/[(1-\gamma)\pi])^2$ as $\Delta \to \pi$. The latter payoff is larger because $\theta_r > \theta_m$ by construction and $\gamma \in [0, 1]$. Therefore, the organisation chooses to form an exclusive club in every period for all p^0 and $\mu \in [0, 1)$.

We have established $p^t \to \bar{p} > 0$. By Lemma 7, $\bar{p}$ is increasing in Δ, and strictly so whenever $\bar{p} < 1$. □

Proof of Proposition 8. To prove the first statement, note that, by Lemma 2(i), if

$$\Delta < \pi - \frac{\theta_r^2}{4\pi\gamma(1-\gamma)},$$

or, equivalently,

$$\theta_r < 2\sqrt{\gamma(1-\gamma)\pi(\pi - \Delta)}, \tag{A15}$$

then p^t is strictly decreasing for all t and moderation occurs.

By inspection, the right-hand side of (A15) is strictly concave in γ and maximised at $\gamma = 1/2$. Evaluating (A15) at $\gamma = 1/2$ yields

$$\theta_r < \sqrt{\pi(\pi - \Delta)}.$$

In addition, (A15) is violated for $\gamma = 0$ and $\gamma = 1$. Hence, by continuity of the right-hand side of (A15) with respect to γ, if $\theta_r < \sqrt{\pi(\pi - \Delta)}$, there exist thresholds $\underline{\gamma} \in (0, \frac{1}{2})$ and $\bar{\gamma} \in (\frac{1}{2}, 1)$ such that (A15) holds if and only if $\gamma \in (\underline{\gamma}, \bar{\gamma})$. The first statement of part (*i*) follows immediately. Recall that $\underline{\gamma}$ and $\bar{\gamma}$ are respectively given by the smallest and largest γ that solve

$$\theta_r = 2\sqrt{\gamma(1-\gamma)\pi(\pi - \Delta)},$$

which is well defined for $\theta_r < \sqrt{\pi(\pi - \Delta)}$. By the implicit function theorem, for $\gamma \in \{\underline{\gamma}, \bar{\gamma}\}$,

$$\frac{\partial \gamma}{\partial \pi} = -\frac{\bar{\gamma}(1-\bar{\gamma})(2\pi - \Delta)}{(1-2\bar{\gamma})\pi(\pi - \Delta)}.$$

As $\underline{\gamma} < \frac{1}{2}$, $\partial\underline{\gamma}/\partial\pi < 0$ and, as $\bar{\gamma} > \frac{1}{2}$, $\partial\bar{\gamma}/\partial\pi > 0$.

Now consider part (*ii*). If $\theta_r \geq \sqrt{\pi(\pi - \Delta)}$ or $\gamma \notin (\underline{\gamma}, \bar{\gamma})$ then (A15) is violated, so Lemma 2(*ii*) applies. Hence, radicalism is possible and occurs under the conditions described in Propositions 2 and 3. □

Proof of Proposition 9. Suppose that $\Delta > \pi - \theta_r^2/(4\gamma(1-\gamma)\pi)$, so $\underline{p}(\alpha)$ and $\bar{p}(\alpha)$ are well defined for $\alpha > 0$. For an exclusive club,

$$p^{t+1} = p^t(\bar{s}(p^t) + (1-\bar{s}(p^t))(1-\gamma)(\alpha + (1-\alpha)p^t)) + (1-p^t)(1-\gamma)(\alpha + (1-\alpha)p^t),$$

so

$$\frac{p^{t+1} - p^t}{p^t} = \bar{s}(p^t)(1 - (1-\gamma)(\alpha + (1-\alpha)p^t)) - \gamma + \alpha(1-\gamma)\frac{1-p^t}{p^t}. \tag{A16}$$

We first show that $\underline{p}(\alpha)$ and $\bar{p}(\alpha)$ are respectively decreasing and increasing in α, strictly so when interior. Recall that $\underline{p}(\alpha)$ and $\bar{p}(\alpha)$ are defined by the smallest and largest p such that the right-hand side of (A16) is non-negative. Differentiating the right-hand side of (A16) with respect to α and evaluating at an arbitrary p yields

$$(1-\gamma)(1 - p\bar{s}(p))\frac{1-p}{p} \geq 0,$$

strictly so for $\bar{s}(p^t) < 1$. Therefore, the right-hand side of (A16) is increasing in α, so $\underline{p}(\alpha)$ is decreasing in α and $\bar{p}(\alpha)$ is increasing in α, both strictly so when interior.

We now show that $\bar{\theta}(\alpha)$ is non-decreasing in α. Recall from the proof of Proposition 3 that

$$\bar{\theta}(\alpha) \equiv \theta_r \min\left\{\frac{\sqrt{p^0}(\pi + \gamma p^0\Delta)}{\pi - \Delta + (1-\gamma)p^0\Delta}, \frac{\pi\,\bar{p}(\alpha)}{\pi - \Delta + (1-\gamma)\bar{p}(\alpha)\Delta}\right\}.$$

The first term is unaffected by α. As we have shown that $\bar{p}(\alpha)$ is strictly increasing in α and the second term is strictly increasing in $\bar{p}(\alpha)$ (by Lemma 5), $\bar{\theta}(\alpha)$ is non-decreasing in α.

We now show that, for sufficiently small γ and sufficiently large α, the right-hand side of (A16) is strictly positive for all $p < \bar{p}(\alpha)$, which implies that $\underline{p}(\alpha) = 0$.

Differentiating the right-hand side of (A16) with respect to p yields

$$\frac{\partial \bar{s}(p)}{\partial p}(1-(1-\gamma)(\alpha+(1-\alpha)p))-\bar{s}(p)(1-\gamma)(1-\alpha)-\frac{\alpha(1-\gamma)}{p^2}.$$

Evaluating the above at $\alpha = 1$ yields

$$\gamma\frac{\partial \bar{s}(p)}{\partial p}-\frac{1-\gamma}{p^2}.$$

For sufficiently small γ, the above is strictly negative, so when an exclusive club is formed, $p^{t+1} - p^t$ is strictly decreasing in p for all $p \in (0, 1)$. Moreover,

$$\lim_{p\to 0}\bar{s}(p)(1-(1-\gamma)(\alpha+(1-\alpha)p))-\gamma+\alpha(1-\gamma)\frac{1-p}{p}>0$$

for all $\alpha > 0$ and

$$\lim_{p\to 1}\bar{s}(p)(1-(1-\gamma)(\alpha+(1-\alpha)p))-\gamma+\alpha(1-\gamma)\frac{1-p}{p}=-\gamma(1-\bar{s}(1))\leq 0,$$

so the right-hand side of (A16) is strictly greater than zero at $p = 0$ and non-positive at $p = 1$. Hence, there is exactly one p such that (A16) equals zero and (A16) is strictly positive for all p lower than this critical p, which implies that $\underline{p}(\alpha) = 0$. By continuity and the fact that (A16) is increasing in α, $\underline{p}(\alpha) = 0$ for all sufficiently large α.

Lastly, we show that, when radicalism occurs for $\alpha = 0$, it occurs for all $\alpha > 0$. By Proposition 3, radicalism occurs for sufficiently large μ, $p^0 > \underline{p}(0)$ and $\theta_m < \bar{\theta}(0)$. We have already shown that if $\theta_m < \bar{\theta}(0)$ then $\theta_m < \bar{\theta}(\alpha)$ for all $\alpha > 0$. In addition, $\underline{p}(\alpha)$ is strictly decreasing in α when interior and is interior at $\alpha = 0$, so radicalism occurs for some $p^0 < \underline{p}(0)$. □

Proof of Proposition 10. (*i*) Let us begin by showing that the organisation can produce radicalisation from all initial states $p^0 \geq \underline{P}$, where $\underline{P} < \underline{p}$.

Fix $p^0 < \underline{p}$. Suppose that an exclusive club forms in every period. By (A1),

$$p^1 - p^0 = p^0[\bar{s}(p^0, \delta^0)(1-(1-\gamma)p^0)-\gamma],$$

which is negative whenever $p^0 < \underline{p}$ by the construction of $\underline{p}$ (when $\delta^0 = 0$).

Now, note that

$$\delta^1 = \bar{s}(p^0, \delta^0)[(1-\sigma)p^0+\sigma] > 0.$$

As $\delta^0 = 0$ by assumption, $\delta^1 > \delta^0$.

It is straightforward to show that $\bar{s}(p, \delta)$ is strictly increasing in δ. Therefore,

$$\begin{aligned} p^2 - p^1 &= p^1[\bar{s}(p^1, \delta^1)(1-(1-\gamma)p^1)-\gamma] \\ &> p^1[\bar{s}(p^1, \delta^0)(1-(1-\gamma)p^1)-\gamma]. \end{aligned}$$

Taking $p^0 \to \underline{p}$, so that $p^1 \to p^0$, (A17) becomes

$$p^2 - p^0 > p^0[\bar{s}(p^0, \delta^0)(1-(1-\gamma)p^0)-\gamma] = 0.$$

Hence, by continuity of $\bar{s}(p^0, 0)$, $p^2 > p^0$ for p^0 sufficiently close to (but less than) $\underline{p}$. Combined with $\delta^2 > 0 = \delta^0$, we have

$$\begin{aligned} p^3 - p^2 &= p^2[\bar{s}(p^2, \delta^2)(1 - (1-\gamma)p^2) - \gamma] \\ &> p^0[\bar{s}(p^0, \delta^0)(1 - (1-\gamma)p^0) - \gamma]. \end{aligned} \tag{A17}$$

Again, the last expression goes to zero, as $p^0 \to \underline{p}$. Hence, for p^0 sufficiently close to $\underline{p}$, $p^3 > p^2$.

As δ^t is strictly increasing in p^{t-1} and $p^2 > p^1$, $\delta^3 > \delta^2$. Combined with $p^3 > p^2$, we have $p^4 > p^3$.

Iterating this argument yields an increasing sequence $\{p^t\}_{t=T}^{\infty}$ from $T = 3$ (i.e., radicalisation), beginning in any initial state p^0 close to, but less than, $\underline{p}$. Therefore, $\underline{P} < \underline{p}$.

By arguments used in Lemma 2, radicalism also occurs for all $p^0 \geq \underline{p}$ when an exclusive club is formed in every period.

We now know that the organisation can produce radicalisation from $p^0 \geq \underline{P}$ by forming an exclusive club. Next, we show that the organisation will choose to do so if μ and θ_r are sufficiently large. Let δ_m denote the limiting δ^t under inclusivity. That is, δ_m is given by the δ that satisfies

$$\underbrace{\left(\frac{\theta_m}{w-\delta}\right)^2}_{\underline{s}(0,\delta)} = \delta.$$

Let δ_r denote the limiting δ^t under exclusivity, with the latter given by the δ that solves

$$\underbrace{\left(\frac{\theta_r\sqrt{\bar{P}}}{w-\delta-\Delta+(1-\gamma)\bar{P}\Delta}\right)^2}_{\bar{s}(\bar{P},\delta)}((1-\sigma)\bar{P}+\sigma) = \delta. \tag{A18}$$

For radicalisation to occur for sufficiently large μ, the analogues of (A9) and (A10) must be satisfied. The analogue of (A9) is

$$\frac{\theta_m}{w-\delta^t+\gamma p^t\Delta} < \frac{\theta_r\sqrt{p^t}}{w-\delta^t-\Delta+(1-\gamma)p^t\Delta}$$

for each t in which screening occurs, which can be rewritten as

$$\frac{\theta_m}{\theta_r} < \frac{\sqrt{p^t}(w-\delta^t+\gamma p^t\Delta)}{w-\delta^t-\Delta+(1-\gamma)p^t\Delta}. \tag{A19}$$

Differentiating the right-hand side of (A19) with respect to δ^t yields

$$\frac{\Delta(1+\gamma p^t-(1-\gamma)p^t)\sqrt{p^t}}{(w-\delta^t-\Delta+(1-\gamma)p^t\Delta)^2},$$

which is strictly positive for all p^t. Hence, it is sufficient for (A19) to hold at $t = 0$ as in the case of no discrimination, which it does for sufficiently large θ_r.

The analogue of (A10) is

$$\frac{\theta_m}{w-\delta_m} < \frac{\theta_r\bar{P}}{w-\delta_r-\Delta+(1-\gamma)\bar{P}\Delta},$$

which can be rewritten as

$$\frac{\theta_m}{\theta_r} \times \frac{w - \delta_r - \Delta(1 - (1 - \gamma)\bar{P})}{w - \delta_m} < \bar{P}. \tag{A20}$$

Recall from the proof of Proposition 4 that $\bar{s}(\bar{p}, \delta) = 1$ if and only if $\bar{p} = 1$. Hence, by (A18), if $\bar{P} = 1$ (and thus $\bar{s}(\bar{P}, \delta) = 1$) then $\delta_r = 1$. Therefore, evaluating (A20) at $\bar{P} = 1$ yields

$$\frac{\theta_m}{\theta_r} \times \frac{w - 1 - \gamma\Delta}{w - \delta_m} < 1. \tag{A21}$$

As $\theta_m < w - 1$ by assumption, $\delta_m < 1$. Given $\delta_m < 1$ and $\theta_m < \theta_r$, (A21) is satisfied. Hence, (A20) is satisfied at $\bar{P} = 1$ (and, by continuity, close to one).

In Proposition 4, we established that $\bar{p} = 1$ if and only if $\theta_r \geq \pi - \gamma\Delta$. Analogously, under endogenous discrimination, $\bar{P} = 1$ if and only if $\theta_r \geq w - 1 - \gamma\Delta$. Hence, (A20) is satisfied for sufficiently large θ_r. This establishes part (*i*).

(*ii*) When radicalism occurs, the sequence $\{p^t\}_{t=T}^{\infty}$ converges to $\bar{P}$, which is the value of p that solves

$$\bar{s}(p, \delta(p))\,(1 - (1 - \gamma)p) = \gamma. \tag{A22}$$

Recall that without discrimination $\bar{p}$ solves

$$\bar{s}(p, 0)\,(1 - (1 - \gamma)p) = \gamma. \tag{A23}$$

We know that if $\bar{s}(\bar{p}, 0) = 1$ then $\bar{p} = 1$ (i.e., the upper bound). Thus, restrict attention to $\bar{s}(\bar{p}, 0) < 1$. In this case, we know that $\bar{s}(\bar{p}, \delta(\bar{p})) > \bar{s}(\bar{p}, 0)$ for $\delta(\bar{p}) > 0$. In addition, by (7), $\delta(p) > 0$ for all $p > 0$. Hence, if (A23) holds at $p = \bar{p}$ then the left-hand side of (A22) is larger than the right-hand side at $p = \bar{p}$.

At $\bar{P}$, the right-hand side of (A22) is strictly decreasing in p. Hence, p must rise above $\bar{p}$ to equate the two sides of (A22). That is, $\bar{P} > \bar{p}$.

(*iii*) Raising σ raises $\delta(p)$ for all $p < 1$ when an exclusive club is formed (see (7)). Hence, we can let $p^1 - p^0$ be more negative and still get $\delta^1 > \delta^0$, as required in the argument in part (*i*). This implies that $\underline{P}$ is strictly decreasing in σ.

Because raising σ raises $\delta(p)$ for all $p < 1$, the same argument used in part (*ii*) implies that, whenever less than one, $\bar{P}$ is strictly increasing in σ. □

University of Oxford, UK
Clarkson University, USA

Additional Supporting Information may be found in the online version of this article:

Replication Package

References

Adida, C.L., Laitin, D.D. and Valfort, M.A. (2016). *Why Muslim Integration Fails in Christian-Heritage Societies*, Cambridge, MA: Harvard University Press.

Aimone, J.A., Iannaccone, L.R., Makowsky, M.D. and Rubin, J. (2013). 'Endogenous group formation via unproductive costs', *Review of Economic Studies*, vol. 80(4), pp. 1215–36.

Akerlof, G.A. and Kranton, R.E. (2000). 'Economics and identity', *Quarterly Journal of Economics*, vol. 115(3), pp. 715–53.

Akerlof, G.A. and Kranton, R.E. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton, NJ: Princeton University Press.

Akerlof, R. (2017). 'Value formation: The role of esteem', *Games and Economic Behavior*, vol. 102, pp. 1–19.

Al-Ajarma, K. (2021). 'After Hajj: Muslim pilgrims refashioning themselves', *Religions*, vol. 12(36), pp. 1–14.

Alesina, A., Carlana, M., Ferrara, E.L. and Pinotti, P. (2018). 'Revealing stereotypes: Evidence from immigrants in schools', Working paper, National Bureau of Economic Research.

Algan, Y., Bisin, A., Manning, A. and Verdier, T. (2012). *Cultural Integration of Immigrants in Europe*, New York: Oxford University Press.

Almagro, M. and Andrés-Cerezo, D. (2020). 'The construction of national identities', *Theoretical Economics*, vol. 15(2), pp. 763–810.

Aoki, K. (1986). 'A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans', *Proceedings of the National Academy of Sciences*, vol. 83(9), pp. 2929–33.

Appiah, K.A. (2006). *Cosmopolitanism: Ethics in a World of Strangers*, New York: W.W. Norton.

Arzheimer, K. and Berning, C.C. (2019). 'How the alternative for Germany (AfD) and their voters veered to the radical right, 2013–2017', *Electoral Studies*, vol. 60, p. 102040.

Aumann, R. (1959). 'Acceptable points in general cooperative n-person games', *Annals of Mathematics Studies*, vol. 40, pp. 287–324.

Auriol, E. and Platteau, J.P. (2017). 'Religious co-option in autocracy: A theory inspired by history', *Journal of Development Economics*, vol. 127, pp. 395–412.

Bénabou, R. and Tirole, J. (2011). 'Identity, morals, and taboos: Beliefs as assets', *Quarterly Journal of Economics*, vol. 126(2), pp. 805–55.

Benmelech, E. and Klor, E.F. (2020). 'What explains the flow of foreign fighters to ISIS?', *Terrorism and Political Violence*, vol. 34(7), pp. 1458–81.

Berger, J.M. (2018). *Extremism*, Cambridge, MA: MIT Press.

Berman, E. (2000). 'Sect, subsidy, and sacrifice: An economist's view of Ultra-Orthodox Jews', *Quarterly Journal of Economics*, vol. 115(3), pp. 905–53.

Berman, E. (2009). *Radical, Religious, and Violent: The New Economics of Terrorism*, Cambridge, MA: MIT Press.

Berman, E., Shapiro, J.N. and Felter, J.H. (2011). 'Can hearts and minds be bought? The economics of counterinsurgency in Iraq', *Journal of Political Economy*, vol. 119(4), pp. 766–819.

Binzel, C. and Carvalho, J.P. (2017). 'Education, social mobility, and religious movements: The Islamic revival in Egypt', Economic Journal, vol. 127(607), pp. 2553–80.

Bisin, A., Patacchini, E., Verdier, T. and Zenou, Y. (2011a). 'Ethnic identity and labour market outcomes of immigrants in Europe', *Economic Policy*, vol. 26(65), pp. 57–92.

Bisin, A., Patacchini, E., Verdier, T. and Zenou, Y. (2011b). 'Formation and persistence of oppositional identities', *European Economic Review*, vol. 55(8), pp. 1046–71.

Bisin, A. and Verdier, T. (2000). ' "Beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits', *Quarterly Journal of Economics*, vol. 115(3), pp. 955–88.

Bisin, A. and Verdier, T. (2001). 'The economics of cultural transmission and the dynamics of preferences', *Journal of Economic Theory*, vol. 97(2), pp. 298–319.

Bisin, A. and Verdier, T. (2017). 'On the joint dynamics of culture and institutions', Working Paper 23375, National Bureau of Economic Research.

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016). 'Stereotypes', *Quarterly Journal of Economics*, vol. 131(4), pp. 1753–94.

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2019). 'Beliefs about gender', *American Economic Review*, vol. 109(3), pp. 739–73.

Boyd, R. and Richerson, P.J. (1988). *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.

Bueno de Mesquita, E. (2005). 'The quality of terror', *American Journal of Political Science*, vol. 49(3), pp. 515–30.

Bueno de Mesquita, E. and Dickson, E.S. (2007). 'The propaganda of the deed: Terrorism, counterterrorism, and mobilization', *American Journal of Political Science*, vol. 51(2), pp. 364–81.

Burleigh, M. (2000). *The Third Reich: A New History*, New York: Hill & Wang.

Carvalho, J.P. (2013). 'Veiling', *Quarterly Journal of Economics*, vol. 128(1), pp. 337–70.

Carvalho, J.P. (2016). 'Identity-based organizations', *American Economic Review: Papers and Proceedings*, vol. 106(5), pp. 410–4.

Carvalho, J.P. (2019). 'Religious clubs: The strategic role of religious identity', in (J.P. Carvalho, S. Iyer and J. Rubin, eds.), *Advances in the Economics of Religion*, pp. 25–41, Cham: Palgrave Macmillan.

Carvalho, J.P. (2020). 'Sacrifice and sorting in clubs', *Forum for Social Economics*, vol. 49(4), pp. 357–69.

Carvalho, J.P. (forthcoming). 'Religion and terrorism: The religious utility hypothesis', in (A. Basuchoudhary and G. Schulze, eds.), *Handbook of the Economics of Terrorism*, Cambridge: Cambridge University Press.

Carvalho, J.P. and Koyama, M. (2016). 'Jewish emancipation and schism: Economic development and religious change', *Journal of Comparative Economics*, vol. 44(3), pp. 562–84.

Carvalho, J.P., Koyama, M. and Sacks, M. (2017). 'Education, identity, and community: Lessons from Jewish emancipation', *Public Choice*, vol. 171(1–2), pp. 119–43.

Carvalho, J.P., Koyama, M. and Williams, C. (2022). 'Resisting education', Working paper, University of Oxford, Department of Economics Discussion Paper 982.

Carvalho, J.P. and Sacks, M. (2021). 'The economics of religious communities', *Journal of Public Economics*, vol. 201, 104481.

Cavalli-Sforza, L.L. and Feldman, M.W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*, vol. 16, Princeton, NJ: Princeton University Press.

Chen, T., McBride, M. and Short, M.B. (2019). 'Dynamics of religious group growth and survival', *Journal for the Scientific Study of Religion*, vol. 58(1), pp. 67–92.

Chen, Y. and Li, S.X. (2009). 'Group identity and social preferences', *American Economic Review*, vol. 99(1), pp. 431–57.

Chudek, M., Heller, S., Birch, S. and Henrich, J. (2012). 'Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning', *Evolution and Human Behavior*, vol. 33(1), pp. 46–56.

Clark, J.A. (2004). *Islam, Charity, and Activism: Middle-class Networks and Social Welfare in Egypt, Jordan, and Yemen*, Bloomington, IN: Indiana University Press.

Connor, P. (2010). 'Contexts of immigrant receptivity and immigrant religious outcomes: The case of Muslims in Western Europe', *Ethnic and Racial Studies*, vol. 33(3), pp. 376–403.

Corno, L., La Ferrara, E. and Burns, J. (2022). 'Interaction, stereotypes and performance: Evidence from South Africa', *American Economic Review*, vol. 112(12), pp. 3848–75.

Dal Bo, E., Finan, F., Folke, O., Persson, T. and Rickne, J. (2023). 'Economic and social outsiders but political insiders: Sweden's populist radical right', *Review of Economic Studies*, vol. 90(2), pp. 675–706.

Douglas, M. (1966). *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*, London: Routledge.

Fan, J., Friedman, D., Gair, J., Iyer, S., Redlicki, B. and Velu, C. (2021). 'A simulation study of how religious fundamentalism takes root', *Journal of Economic Behavior and Organization*, vol. 192, pp. 465–81.

Forstenlechner, I. and Al-Waqfi, M.A. (2010). ' "A job interview for Mo, but none for Mohammed": Religious discrimination against immigrants in Austria and Germany', *Personnel Review*, vol. 39(6), pp. 767–84.

Fouka, V. (2020). 'Backlash: The unintended effects of language prohibition in US schools after World War I', *Review of Economic Studies*, vol. 87(1), pp. 204–39.

Fukuyama, F. (2018). *Identity: Contemporary Identity Politics and the Struggle for Recognition*, London: Profile Books.

Gaus, G. (2019). *The Tyranny of the Ideal: Justice in a Diverse Society*, Princeton, NJ: Princeton University Press.

Genicot, G. (2022). 'Tolerance and compromise in social networks', *Journal of Political Economy*, vol. 130(1), pp. 94–120.

Government of Canada. (1988). 'Canadian Multiculturalism Act', https://laws-lois.justice.gc.ca/eng/acts/c-18.7/page-1.html.

Government of Denmark. (2018). 'One Denmark without parallel societies: No ghettos in 2030', https://www.regeringen.dk/media/4937/publikation_%C3%A9t-danmark-uden-parallelsamfund.pdf.

Gulis, G., Safi, M. and Linde, D. (2022). 'Rapid health impact assessment of a Danish policy document: One Denmark without parallel societies: No ghettos in 2030', *Journal of Public Health*, vol. 30, pp. 983–9.

Han, B., Hirshleifer, D. and Walden, J. (2021). 'Social transmission bias and investor behavior', *Journal of Financial and Quantitative Analysis*, vol. 57(1), pp. 390–412.

Hanes, E. and Machin, S. (2014). 'Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11', *Journal of Contemporary Criminal Justice*, vol. 30(3), pp. 247–67.

Hauk, E. and Mueller, H. (2015). 'Cultural leaders and the clash of civilizations', *Journal of Conflict Resolution*, vol. 59(3), pp. 367–400.

Henrich, J. and Gil-White, F.J. (2001). 'The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission', *Evolution and Human Behavior*, vol. 22(3), pp. 165–96.

Hirschkind, C. (2006). *The Ethical Soundscape: Cassette Sermons and Islamic Counterpublics*, New York: Columbia University Press.

Hume, D. (1983). *The History of England: From the Invasion of Julius Caesar to the Revolution in 1688*, Indianapolis, IN: Liberty Fund.

Iannaccone, L.R. (1992). 'Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives', *Journal of Political Economy*, vol. 100(2), pp. 271–91.

Iannaccone, L.R. and Berman, E. (2006). 'Religious extremism: The good, the bad, and the deadly', *Public Choice*, vol. 128(1–2), pp. 109–29.

Ibrahim, S.E. (1980). 'Anatomy of Egypt's militant Islamic groups: Methodological note and preliminary findings', *International Journal of Middle East Studies*, vol. 12(4), pp. 423–53.

Iyigun, M., Rubin, J. and Seror, A. (2021). 'A theory of cultural revivals', *European Economic Review*, vol. 135, 103734.

Kahneman, D. and Tversky, A. (1972). 'Subjective probability: A judgment of representativeness', *Cognitive Psychology*, vol. 3(3), pp. 430–54.

Kepel, G. (2003). *Muslim Extremism in Egypt: The Prophet and Pharaoh*, Berkeley, CA: University of California Press.

Kepel, G. (2004). *The War for Muslim Minds: Islam and the West*, Cambridge, MA: Belknap Press of Harvard University Press.

Kirk, A. (2016). 'Iraq and Syria: How many foreign fighters are fighting for Isil?', *The Telegraph*, 24 March, https://www.telegraph.co.uk/news/2016/03/29/iraq-and-syria-how-many-foreign-fighters-are-fighting-for-isil.

Kukathas, C. (2003). *The Liberal Archipelago: A Theory of Diversity and Freedom*, New York: Oxford University Press.

Kuran, T. (2004). *Islam and Mammon*, Princeton, NJ: Princeton University Press.

Kymlicka, W. (1995). *Multicultural Citizenship: A Liberal Theory of Minority Rights*, Oxford: Clarendon Press.
Laland, K.N., Odling-Smee, J. and Feldman, M.W. (2001). 'Cultural niche construction and human evolution', *Journal of Evolutionary Biology*, vol. 14(1), pp. 22–33.
Lepore, C.R. (2012). 'Asserting state sovereignty over national communities of Islam in the United States and Britain: Sharia courts as a tool of Muslim accommodation and integration', *Washington University Global Studies Law Review*, vol. 11, pp. 669–92.
Levy, G. and Razin, R. (2012). 'Religious beliefs, religious participation, and cooperation', *American Economic Journal: Microeconomics*, vol. 4(3), pp. 121–51.
Levy, G. and Razin, R. (2014). 'Rituals or good works: Social signaling in religious organizations', *Journal of the European Economic Association*, vol. 12(5), pp. 1317–60.
Masci, D. (2019). 'Many Americans see religious discrimination in U.S.—especially against Muslims', *Pew Research Center: Fact Tank—News in the Numbers*, 17 May, https://www.pewresearch.org/fact-tank/2019/05/17/many-americans-see-religious-discrimination-in-u-s-especially-against-muslims/
McBride, M. (2008). 'Religious pluralism and religious participation: A game theoretic analysis', *American Journal of Sociology*, vol. 114(1), pp. 77–106.
McBride, M. (2010). 'Religious market competition in a richer world', *Economica*, vol. 77(305), pp. 148–71.
McBride, M. (2015). 'Why churches need free-riders: Religious capital formation and religious group survival', *Journal of Behavioral and Experimental Economics*, vol. 58, pp. 77–87.
Mitchell, R.P. (1993). *The Society of the Muslim Brothers*, New York: Oxford University Press.
Mitts, T. (2019). 'From isolation to radicalization: Anti-Muslim hostility and support for ISIS in the West', *American Political Science Review*, vol. 113(1), pp. 173–94.
Moustafa, T. (2000). 'Conflict and cooperation between the state and religious institutions in contemporary Egypt', *International Journal of Middle East Studies*, vol. 32(1), pp. 3–22.
Muldoon, R. (2016). *Social Contract Theory for a Diverse World: Beyond Tolerance*, New York: Routledge.
Odling-Smee, F.J., Laland, K.N. and Feldman, M.W. (2003). *Niche Construction: The Neglected Process in Evolution*, Princeton, NJ: Princeton University Press.
Packer, G. (2015). 'The other France: Are the suburbs of Paris incubators of terrorism?', *The New Yorker*, 24 August, https://www.newyorker.com/magazine/2015/08/31/the-other-france.
Petraeus, D. (2016). 'Anti-Muslim bigotry aids Islamist terrorists', *Washington Post*, 13 May, https://www.washingtonpost.com/opinions/david-petraeus-anti-muslim-bigotry-aids-islamist-terrorists/2016/05/12/5ab50740-16aa-11e6-924d-838753295f9a_story.html.
Pew-Templeton. (2022). 'Global religious futures project', https://web.archive.org/web/20221023190813/http://www.globalreligiousfutures.org/regions/europe/religious_demography#/?affiliations_religion_id=0&affiliations_year=2010
Prummer, A. and Siedlarek, J.P. (2017). 'Community leaders and the preservation of cultural traits', *Journal of Economic Theory*, vol. 168, pp. 143–76.
Reader, I. (2013). *Religious Violence in Contemporary Japan: The Case of Aum Shinrikyo*, London: Routledge.
Roy, O. (1994). *The Failure of Political Islam*, Cambridge, MA: Harvard University Press.
Roy, O. (2017). *Jihad and Death: The Global Appeal of Islamic State*, London: Hurst.
Rubin, J. (2017). *Rulers, Religion, and Riches: Why the West got Rich and the Middle East Did Not*, Cambridge: Cambridge University Press.
Sen, A. (2006). *Identity and Violence: The Illusion of Destiny*, New York: Norton.
Shayo, M. (2009). 'A model of social identity with an application to political economy: Nation, class, and redistribution', *American Political Science Review*, vol. 103(2), pp. 147–74.
Smith, A. (2003). *The Wealth of Nations*, New York: Random House.
Sundquist, V.H. (2010). 'Political terrorism: An historical case study of the Italian Red Brigades', *Journal of Strategic Security*, vol. 3(3), pp. 53–68.
Verdier, T. and Zenou, Y. (2015). 'The role of cultural leaders in the transmission of preferences', *Economics Letters*, vol. 136, pp. 158–61.
Verdier, T. and Zenou, Y. (2018). 'Cultural leaders and the dynamics of assimilation', *Journal of Economic Theory*, vol. 175, pp. 374–414.
Vidino, L. (2020). *The Closed Circle: Joining and Leaving the Muslim Brotherhood in the West*, New York: Columbia University Press.
Wickham, C.R. (2002). *Mobilizing Islam: Religion, Activism and Political Change in Egypt*, New York: Columbia University Press.
Wickham, C.R. (2015). *The Muslim Brotherhood: Evolution of an Islamist Movement*, Princeton, NJ: Princeton University Press.
Wike, R., Stokes, B. and Simmons, K. (2016). 'Europeans fear wave of refugees will mean more terrorism, fewer jobs', *Pew Research Center: Global Attitudes & Trends*, 11 July, https://www.pewresearch.org/global/2016/07/11/europeans-fear-wave-of-refugees-will-mean-more-terrorism-fewer-jobs.
Wright, S.A. (2014). *Armageddon in Waco: Critical Perspectives on the Branch Davidian Conflict*, Chicago: University of Chicago Press.

The Economic Journal, **134** (*April*), 1019–1068 https://doi.org/10.1093/ej/uead094
Advance Access Publication Date: 26 October 2023