CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/142371/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Whitaker, Roger M. , Colombo, Gualtiero B., Turner, Liam , Dunham, Yarrow, Doyle, Darren K., Roy, Eilish M. and Giammanco, Cheryl A. 2022. The coevolution of social networks and cognitive dissonance. IEEE Transactions on Computational Social Systems 9 (2) , pp. 376-393. 10.1109/TCSS.2021.3090833

Publishers page: http://dx.doi.org/10.1109/TCSS.2021.3090833

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

The Coevolution of Social Networks and Cognitive Dissonance

Roger M. Whitaker[1], Gualtiero B. Colombo[1], Liam Turner[1], Yarrow Dunham[2], Darren K. Doyle[3], Eilish M. Roy[3], and Cheryl A. Giammanco [4]

[1]Crime and Security Research Institute, Cardiff University, Cardiff, UK
[2]Department of Psychology, Yale University, New Haven, USA
[3]Defence Science and Technology Laboratory, Human and Social Sciences Group, Porton Down, Salisbury, UK
[4]Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, Maryland, USA

Cognitive dissonance is well-understood as a significant psychological motivator of behaviour. It can be experienced vicariously when a member of one's social group acts inconsistently to expectations. In this paper we explore the network implications from individuals reconciling cognitive friction when their neighbours hold alternative views. Through agent-based modelling, we introduce a framework to explore the sensitivity of behaviour on social network structure, in response to vicarious dissonance. The model allows us to understand how and why vicarious dissonance may contribute to polarisation, both in terms of network structure and the convictions held by individuals. Alternative response behaviours are each found to be highly effective in reducing the cognitive dissonance felt across a population, but with wide ranging outcomes for the population as a whole. The results highlight the important role of neutrality and tolerance in retaining social cohesion, while showing how easily this can be disrupted. The model presents a useful tool for further research, allowing bespoke scenarios to be investigated.

***Index Terms*—social networks, cognitive dissonance, social influence, agent based modelling**

I. Introduction

The ease with which populations are susceptible to large-scale ideological polarisation has been well seen in numerous global events. The underlying motivations for these behaviours are complex, but the psychological stress experienced by those involved is an important factor in motivating action as a result of conflicting attitudes and beliefs. This represents a form of *cognitive dissonance* - the psychological distress felt through misalignment of ones beliefs and actions with the world around them. In particular, cognitive dissonance can be felt through the social network - when actors with whom strong ties are held act in a manner that is difficult to reconcile with ones own position. This is a form of *vicarious dissonance* and motivates individuals to restore consonance [24]: for example, responses may involve individuals changing their own views or the relationships they hold with others. Collectively, these actions can aggregate and cascade across a group or population, invoking cognitive dissonance for others, resulting in further response cycles. While cognitive dissonance has been studied at the individual level, the implications for a population's social network structure have not been extensively investigated.

The focus of this paper is the interplay between cognitive dissonance and social networks [58] and, in particular, on situations in which individuals are not themselves engaging in socially incompatible behaviours but rather observing and detecting inconsistent actions from other group members. We specifically explore the network implications from individuals seeking to reconcile friction induced from holding divergent views to their neighbours. Action in response to cognitive dissonance may disrupt the network structure, and this is a function of the actor's sensitivity to tolerating alternative views. These dynamics are relevant to the establishment and sustenance of both formal and informal ideological groups, where members identify with common beliefs or views (e.g., politics, religion) as well as being relevant to how individual relationships are maintained.

We model social network evolution as a consequence of the vicarious dissonance caused by an agent's neighbours. We use a single dimensional proposition, representing for instance a single strong belief or a commitment to a full ideology, against which all individuals are assumed to hold a *conviction* level, analogous to an opinion held on a continuous scale of strength. This encompasses an individual's attitude strength. The extent to which an individual can *tolerate* deviation from their own conviction without experiencing disruptive cognitive dissonance, is central to our model. Using conviction and tolerance, we investigate key fundamental behaviours triggered in response to different levels of cognitive dissonance [57]. These concern: severing relationships with others, seeking to persuade others of one's own conviction, or assimilating with others conviction levels. Social influence [56] plays an important role in this context, with neighbours collectively providing the basis for cognitive dissonance to accumulate. We explore how these factors affect the structure of the social network as agents in the social network pursue cognitive dissonance reduction. In particular we examine the interplay between thresholds of cognitive dissonance that trigger different behavioural responses, and consider the extent to which individual differences in reconciling cognitive dissonance affect network evolution.

II. Related Literature

A. Cognitive Dissonance

Cognitive dissonance refers to a state of discomfort that results from experiencing incompatible cognitions. In an interpersonal setting, this represents the discomfort experienced

due to inconsistency between attitudes and behaviour [58], which individuals are driven to diminish. Judgment of the incompatibility may arise from social norms or expectations [75], which the individual may acquire through a group or culture [57]. Insights into cognitive dissonance date back to the classical dissonance theory of Festinger [25] and Heider's balance theory [37], [38], which sought to formalise and potentially counter cognitive inconsistencies that an individual may experience through a network of relationships.

As noted by Matz et al [57], Festinger originally proposed that dissonance emanated from four sources, with three of these (the consequences of decisions, forced compliance, exposure to information) representing individual level processes. The forth phenomena was the social group, representing both a source of cognitive dissonance and a vehicle for its reduction. So-called *vicarious dissonance* [68] involves dissonance being felt through relationships defined in groups where individuals derive a component of their identity. A group member can only restore consonance through variables under their individual control - which may lead the individual to realign their own attitude to accept the originally dissonant proposition, particularly if that is being presented by an influential group member.

Engaging in individual attitude change to assimilate with others is noted [68] as potentially an effective means of coping with the inconsistency of others. Matz et al [57] consider disagreement in a group context and find that a range of interpersonal strategies mitigate dissonance - including changing one's own position or joining an attitudinally congenial group, as well as seeking to influence the opposing attitudes. Smith et al [74] also find greater conformity to in-group norms under high levels of uncertainty. These phenomena highlight the potential power of groups upon the individual, with relationships involving strong personal identification being more likely to mediate experiences [38], [47], [40]. Furthermore, it is likely that multiple relationships have a combined effect on an individual's cognitive dissonance, as repeated exposure to others provides a form of complex contagion, supporting confirmation bias [55].

B. Social Influence

Social influence arises from the interaction between multiple sources and multiple targets of influence over time [56]. This is highly relevant to social cognition, affecting group processes, social power and attitudes. Social influence can involve interactions that may both cause and mitigate cognitive dissonance for the individual [57], resulting in changes at the collective level (e.g., group or network) when many individuals are involved. Simple interactions modelled between individuals can aggregate to complex and dynamic phenomena, such as polarisation or convergence of different forms of opinion [26], [1]. However, relatively little explicit attention has been paid to the role of vicarious dissonance in this context (e.g., [41], [3]). Social influence is also frequently considered in tandem with homophily (e.g., [27], [32]). Homophily [61] represents the tendency for attraction between similar individuals and its effects are well-documented in diverse literature (e.g., [66], [42], [16]), leading to clustering effects.

Fundamental understanding of this issue has come from Axelrod's model of cultural evolution [5]. Here, culture is considered as a form of opinion, being a vector of discrete features, that are held by each individual. Axelrod's approach combines influence and homophily, with similarity of the cultures held between pairs of neighbours promoting the further copying of traits, leading to regions of shared culture.

While Axelrod [5] assumes that culture is a simple contagion that is conveyed through dyadic interactions, many aspects of culture are only adopted after reinforcement from multiple sources (e.g., [15]). Accordingly, other extensions to Axelrod's approach have emerged (e.g., [27], [17]), including alignment with social impact theory [48], which proposes a cumulative but increasingly marginal effect of influence from additional relationships. Such social impact is proposed as a function of the strength, immediacy and number of people that convey the impact, and has been expressed in numerous computational forms (e.g., [65], [69], [46]).

Limited contributions explicitly address cognitive dissonance or vicarious dissonance in relation to social influence. Crano [19] proposes a context and comparison based model to address cognitive dissonance and social influence. This modifies Moscovici's conversion theory [63] between majority and minority in-group members, explaining how group cohesion can be maintained based on tolerance within dialogue and related attitudes. However, the relationship between vicarious dissonance and structural evolution remains without a full understanding.

C. Contribution

By combining responses to dissonance as felt through connectivity with others, our approach is to investigate how vicarious dissonance coevolves with social network structure. Vicarious dissonance is the cognitive dissonance an agent experiences through social network structure as a consequence of reconciling the behaviour of neighbours. The interplay here is well-seen in the wider population, particularly when groups form around conviction to ideological concepts (e.g., [83]) such that interaction with the out-group causes cognitive friction to the in-group. As well as the extent of conviction towards a concept, we consider that cognitive dissonance itself may be subject to social influence: in essence the capacity for cognitive friction being something which may be transmitted as a behavioural contagion, alongside the extent of attitude strength. This may occur in the context of group identity, and is therefore relevant in cases when individual identity fusion is strong. We present the model in the following section, and relate it to the wider literature.

III. Model and Justification

We seek to model how individual attitudes and behaviour may be influenced in the context of a social network due to vicarious dissonance. Specifically agents may act to reduce the cognitive dissonance emanating from the relations held with others who have differing levels of conviction towards a significant issue of common interest. This is based on findings of Matz and Wood [57], [86] who observe three distinct

response behaviours as a consequence of cognitive dissonance. To examine the impact of these individual behaviours, we adopt a social network structure, involving fixed traits carried by individuals independent of their beliefs (Section III-A). Each agent expresses its own *conviction* aligned to attitude strength (Section III-B), and the extent to which an individual may experience cognitive dissonance from alternative views is incorporated through a *tolerance* parameter (Section III-D). We note that attitude strength and tolerance of alternative convictions are likely to be related, and accordingly these factors are correlated (Section III-C). In response to cognitive dissonance, we consider three forms of behaviour [57], namely breaking of relationships, persuasion of others or assimilation with alternative views (Section III-E).

A. Network Representation

We note that individuals may possess traits that underpin commonality - such as geography, heritage, previous experience or other markers that are unlikely to change over time. These provide individuals with a basis to identify with each other and allow homophily to take hold, giving the potential for a relationship between individuals to be more influential, with two individuals feeling that they are to some degree "in-group". Alongside this, individuals may also hold additional attitudinally based characteristics based on their worldview and beliefs, that are held with a particular strength. If "in-group" individuals find these attitudes incongruent to their own, significant vicarious dissonance may be experienced, with an individual struggling to sustain the "in-group" relationship.

To explore this further we apply a graph-based representation $G = (V, E)$ over a population where individual agents form the nodes $i \in V$, and where agents may influence each other through relationships. To represent immutable characteristics, an agent i carries fixed traits that are represented as a sequence $f_1^i, \ldots, f_{n_f}^i$, where n_f is the number of fixed traits that each agent holds. The similarity of agents i and j, denoted sim_{ij}, is based on the proportion of fixed traits held in common, where:

$$sim_{ij} = \frac{|\{k : f_k^i = f_k^j\}|}{n_f}$$

Relationships between individuals are modelled using directed edges: an edge $(i, j) \in E$ indicates that i receives influence from j, and we refer to this as i *following* j. We assume that an individual agent has a fixed capacity to follow others (i.e., fixed out-degree), while being followed by others is unconstrained.

B. Modelling Attitude Strength and Tolerance

We define *conviction* to be an abstract form of attitude strength in relation to a fundamental belief that is relevant to the population, on which all members are assumed to have a view. Each agent i holds a measure of their conviction, c_i, representing direction and extremity, on an integer scale of 0 to 10, with 5 representing neutrality. Conviction reflects personal importance to the individual, and may also reflect the extent the attitude is rooted in one's sense of identity and is therefore interconnected with other held attitudes and beliefs. From following another agent j, an agent i is potentially exposed to alternative levels of conviction that provide a source of influence on j. Specifically, each agent holds a level of tolerance t_i, $(0 \leq t_i \leq 5)$, that characterises the extent that i can tolerate the conviction level of another agent j without experiencing unbearable cognitive dissonance. If i follows j then i can tolerate j's conviction level c_j if

$$c_i - t_{ij}^* \leq c_j \leq c_i + t_{ij}^*$$

where $t_{ij}^* = (1 - sim_{ij})t_i$. In this formulation i's tolerance of j is mediated by the strength of similarity between i and j. When this is high, t_{ij} ensures that i's tolerance of counter attitudinal influence from j is low. In other words, a stronger relationship between i and j, based on similarity, makes it more challenging for i to reconcile alternative attitudes of j, consistent with the literature [57].

C. The Relationship Between Conviction and Tolerance

Because conviction encompasses the extent to which an individual's attitude is consequential, it exerts strength over thinking and behaviour. Therefore high levels of conviction can correlate with individual certainty, which in many cases (but not all) may lead an individual to discount alternative views. To account for this we relate each particular conviction level with an expected tolerance level. If an agent i holds a conviction level c_i, then its tolerance level, t_i is defined as $t_i = f^T(c_i)$, where f^T represents a probability distribution. We specifically consider f^T as a normal distribution (i.e., extreme conviction results in low tolerance and vice versa) as well as considering f^T as a uniform distribution, which provides a useful baseline against which we can determine the effect of tolerance levels.

D. Assessing Cognitive Dissonance

To establish the extent of cognitive dissonance verses cognitive reinforcement from following someone, we calculate j's *deviation from* i's *tolerance*, denoted d_{ij}, where

$$d_{ij} = t_{ij}^* - |c_i - c_j|.$$

If d_{ij} is zero or positive, then c_j is within i's tolerance range, and outside of this range otherwise. Assuming the agents that i follows are denoted by N_i, then d^i, the total deviation from tolerance encountered by i, while mediating for strength of relationships due to similarity, is

$$d^i = \sum_{j \in N_i} d_{ij} \cdot sim_{ij}$$

This provides an overall measure of the balance of combined attitudes across i's neighbourhood, also taking into account the strength of convictions as mediated by the similarity between agents. It also generalises the classical notion of balance in triadic relationships [41], with d^i indicating the extent to which i faces self-consistency across its neighbourhood. Note that this approach assumes that the strength of influence is mediated by similarity between individuals, aligned to

homophilic attraction. For large neighbourhoods, this could be generalised further, to also include mediation by the number of links sustained.

We define the cognitive dissonance for agent i as:

$$cd^i = -d^i$$

which allows us to discuss the magnitude of a negative quantity (i.e., cognitive dissonance) without confusion.

E. Responding to Cognitive Dissonance

To invoke population dynamics from vicarious dissonance, agents are randomly called to assess their level of cognitive dissonance (cd^i) due to neighbours (Section III-D), which invokes a potential response. We use the framework of Matz and Wood [57] in formulating the actions in response to the extent of cognitive dissonance. We introduce three thresholds, $T_0 < T_1 < T_2$, that categorise when different levels of behaviour are triggered in response to cognitive dissonance. The thresholds align with increased levels of cognitive dissonance. If $cd^i \leq T_0$ then agent i is defined as being able to manage its level of dissonance, if any, and i does not act, other than a small chance ($m = 0.001$) that a mutation takes place through randomly selecting a neighbour, breaking the link and rewiring to a random alternative agent. We refer to not acting as the *"do nothing"* response behaviour.

If $cd^i > T_0$ then agent i applies the following procedures to countenance dissonance, where we let N_i denote the neighbours of i and $N_i' \subset N_i$ contain those neighbours j that are within i's tolerance; that is $N_i' = \{j \in N_i : t_{ij} \geq |c_i - c_j|\}$. In particular:

- If cognitive dissonance (cd^i) is relatively low, say $T_0 < cd^i \leq T_1$, then i assimilates with its neighbours by updating its own conviction level to potentially reduce dissonance. The choice of a new conviction level for i occurs on a probabilistic basis. The probability of that i chooses conviction level c_j from neighbour $j \in N_i'$ is
$$\frac{d_{ij} + K}{\sum_{j \in N_i'}(d_{ij} + K)}$$
where K is a small constant. We refer to this as the *change self* response behaviour.
- If cognitive dissonance (cd^i) is moderate, say $T_1 < cd^i \leq T_2$, then i seeks to persuade one of its neighbours to change its conviction level. A neighbour $j \in N_i$ is selected, upon which persuasion is performed. Neighbour $j \in N_i$ is selected with probability
$$\frac{sim_{ij} + K}{\sum_{k \in N_i}(sim_{ik} + K)}$$
where K is a small constant. Upon selection, neighbour j is persuaded to adopt i's conviction level c_i with probability sim_{ij}. We refer to this as the *change other* response behaviour.
- If cognitive dissonance (cd^i) is high, say $cd^i > T_2$, then agent i seeks to reduce the total deviation from tolerance by ceasing to follow an agent that is contributing the most cognitive dissonance, namely $j \in N_i$ such that $-d_{ij} \geq -d_{ik}$, for all $k \in N_i$, $k \neq i$. Such an agent j is then removed from N_i, and i follows an alternative agent selected at random from the population, which by definition is added to N_i. We refer to this as the *rewire* response behaviour.

The thresholds T_0, T_1, T_2 govern when and how significantly agents respond to different levels of cognitive dissonance. Therefore we experiment by varying the thresholds T_0, T_1, T_2 that trigger responses to cognitive dissonance, and observe the implications for the population and its interconnection. Note that the ordering of response behaviours in our model follows an assumption of egocentricity. When individual's tolerance reduces and cognitive dissonance increases, the agents work progressively outwards in trying to resolve their cognitive dissonance, changing themselves if cognitive dissonance is perceived to be low (i.e., $T_0 < cd^i \leq T_1$), before seeking to change the views of others if cognitive dissonance is moderate (i.e., $T_1 < cd^i \leq T_2$), before removing links to others (i.e., rewiring) if their cognitive dissonance is significant ($T_2 < cd^i$). This aligns with observed response behaviour towards vicarious hypocrisy [28]. In this case, when the levels of perceived dissonance are rising, changing one's own attitude is the most likely response, while the consequences of breaking a valuable link are perceived as potentially the most damaging to self in terms of self esteem and social acceptance [68], [76].

However, individuals that rely on inner dispositions and are less sensitive to the social consequences of their behaviour as well as those who had an important value violated by the behaviour of the group, such as the agents with extreme convictions and low tolerances in our model, are more inclined to respond to dissonance by either reducing their identification with the group or engaging in pro-value and persuasive behaviour [86], [58], [60].

F. Model Dynamics and Metrics

From an initial starting network configuration, agents are each allocated a sequence of numerical traits at random ($n_f = 5$). Note that each agent's traits remain fixed throughout the simulation. This is intentional, so that we can observe how cognitive dissonance is resolved independent from trait-driven homophilic attraction. Random agents are repeatedly selected from the population V for updating. Each such iteration is called a time step. At each time step, the selected agent i calculates the cognitive dissonance cd^i felt as a consequence of the neighbours N_i that are followed. If $cd^i < T_0$ then i is able to manage the cognitive dissonance that it experiences. Alternatively actions are taken as defined in Section III-E, which are a function of the thresholds T_0, T_1 and T_2. We experiment with uniform threshold values for all agents verses different threshold values for particular subsets of agents. The number of agent selections from the population is fixed to reflect the size of the test problem, and set to ensure convergence is achieved.

To assess the effects of the model on the population we use a range of metrics as follows:

- The average cognitive dissonance, conviction and tolerance levels;

TABLE I
SCENARIOS CONSIDERED WHERE THRESHOLDS ARE USED TO INVOKE DIFFERENT AGENT RESPONSE BEHAVIOURS ACROSS ALL AGENTS

Scenario	Main Response	T_0	T_1	T_2
A	mixed baseline	-3	0	3
B	do nothing	3	3	3
C	change self	-3	3	3
D	change other	-3	-3	3
E	rewire	-3	-3	-3

- The number of distinct tolerance levels present in a given population;
- Assortativity by conviction level, representing the proportion of edges in the network that connect agents with the same conviction;
- Average local diversity in neighbourhood conviction levels (div_{avg}), where an agent i's neighbourhood diversity is defined as $div_i = \sum_{j \in N_i} |c_i - c_j|$ and $div_{avg} = \sum_{k \in V} div_k / |V|$;
- Average shortest path length between all pairs of agents, both for whole network and within the induced subnetwork from each conviction level;
- Average neighbourhood similarity (sim_{avg}) where neighbourhood similarity for agent i is defined as $sim^i_{avg} = \sum_{j \in N_i} sim_{ij} / |N_i|$ and $sim_{avg} = \sum_{k \in V} sim^k_{avg} / |V|$.

IV. RESULTS

We experiment using three alternative starting configurations. Firstly, consistent with the work of Axelrod [5] and subsequent extensions (e.g., [17]), we use a regular lattice of 100 agents, which provides a useful baseline because it can be readily visualised. Secondly, we consider the effects of scaling, adopting a regular lattice of 10,000 agents. Thirdly, to assess the effects of social network structure, we generate scale-free networks according to the Barabasi-Albert model [7] again using 10,000 agents. This is constructed applying a fixed agent out-degree of four (i.e., agents have roughly the same cognitive capacity to follow others) while assuming freedom for an agent's in-degree (i.e., some agents can become highly popular).

In each starting configuration, we assume an initial allocation of 11 discrete conviction levels (0 - 10 inclusive). Throughout, results concern the average from five randomly seeded runs and the number of iterations of a random agent being selected for a behavioural response depends on the scenario. These have been chosen to give a sufficient window for convergence of aggregate cognitive dissonance to be observed, involving 100,000 iterations for the 100-agent lattice, 1 million iterations for the 10,000-agent lattice, and 10 million iterations for the scale-free network. We assume that the relationship between conviction and tolerance is governed by the normal distribution as presented in Figure 1. Importantly, *when an agent changes its conviction level in the simulation, we assume that it changes its tolerance level*. Note that extreme conviction levels align with low tolerance and vice-versa.

We perform experimentation using two sets of scenarios. Firstly in Section IV-A, we consider parameters settings from Table I that invoke each of the different responses to cognitive dissonance (Section III-E), applying these across the whole population, denoted as Scenarios A to E. Parameters for T_0, T_1 and T_2 have been established as minimal values that invoke response behaviours aligned to Section III-E, with a combination of these values (Scenario A) allowing for a mix of response behaviours to take place across the population. This provides a useful baseline (denoted as *mixed baseline*). Results are summarised in Table II. Secondly, in Section IV-B, while invoking the mixed baseline, we consider the consequences of particular sub-populations of agents being impervious to influence. This is defined in Table III and denoted by Scenarios F to H, with results in Table IV. Finally, in Section IV-C, we perform a meta-analysis, to identify potential correspondence to the phenomena that we observe from experimentation.

Note that Figures 2, 4 and 5 present examples of a single network with agent's conviction and tolerance levels represented by colours. This allows the interplay between topology and agent characteristics to be observed. In contrast, Figures 3 and 6 present distributions of the agent's final conviction (c_i), tolerance (t_i) and cognitive dissonance (cd^i) levels, taken from a sample of five runs for each starting network configuration.

A. Effects of Alternative Responses to Cognitive Dissonance

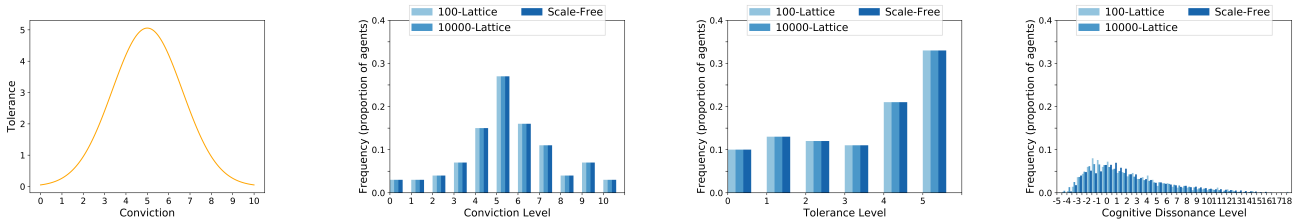
We assume all agents seek to resolve cognitive dissonance through the same triggers (i.e., all agents respond using the same thresholds). Thresholds T_0, T_1 and T_2 are set to promote different response behaviours as shown in Table I.

Scenario A - Mixed Baseline

With the thresholds as defined in Table I, agents act with a combination of responses to counter cognitive dissonance. For example, for the 10,000-agent scale-free network this results in 14.75% of actions being 'do nothing', 32.34% 'change self', 44.70% 'change others' and 8.21% 'rewire'. The results (Table II and visualisation in Figure 2) show that mixed responses are sufficient to significantly reduce overall cognitive dissonance across all possible starting configurations (Figure 3), resulting in aggregate cognitive dissonance levels that are near-zero or negative, on average. The mixed baseline thresholds are sufficient to prevent polarisation, and interestingly, the responses to cognitive dissonance dissolve the scale-free structure, with the scale-free network and 10,000-agent lattice evolving into networks characterised by similar metrics (Table II)). Note that although there is limited rewiring undertaken relative to other actions, this action is sufficient to significantly alter the network topology, reducing the average shortest path lengths for the lattice starting configurations while considerably increasing path lengths given the scale-free network configuration. Figure 3 also indicates that all starting configurations tend to result in a long tail distribution for cognitive dissonance, with peak values in the range [-1,1]. Consistent with this, the associated conviction and tolerance distributions (Figure 3) exhibit clear similarities across the alternative starting configurations.

Scenario B - Do Nothing

The thresholds defined in Table I mainly result in a nil response with only a fractional chance of rewiring due to



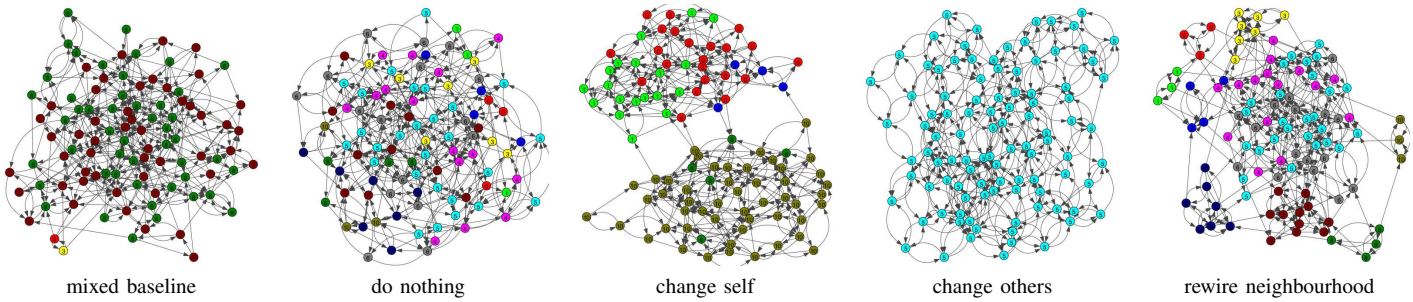
(a) Expected conviction-tolerance relationship

(b) Distribution of agent conviction levels (c_i)

(c) Distribution of agent tolerance levels (t_i)

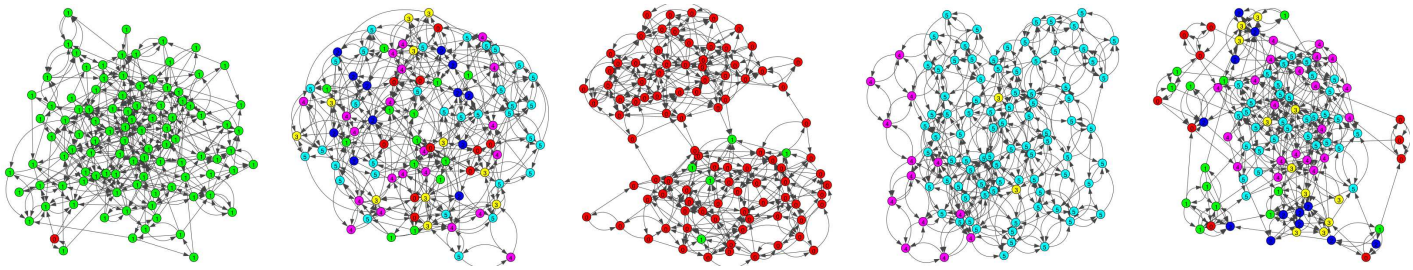
(d) Distribution of agent cognitive dissonance levels (cd^i)

Fig. 1. Expected relationship between an agent's conviction and tolerance level, distributions of conviction (c_i), tolerance (t_i) and the resulting cognitive dissonance (cd^i) at time-step 0 for alternative starting configurations.



mixed baseline | do nothing | change self | change others | rewire neighbourhood

Conviction levels held by agents at the final iteration, after starting from a 100-agent lattice configuration.



mixed baseline | do nothing | change self | change others | rewire neighbourhood

Tolerance levels held by agents at the final iteration, after starting from a 100-agent lattice configuration.

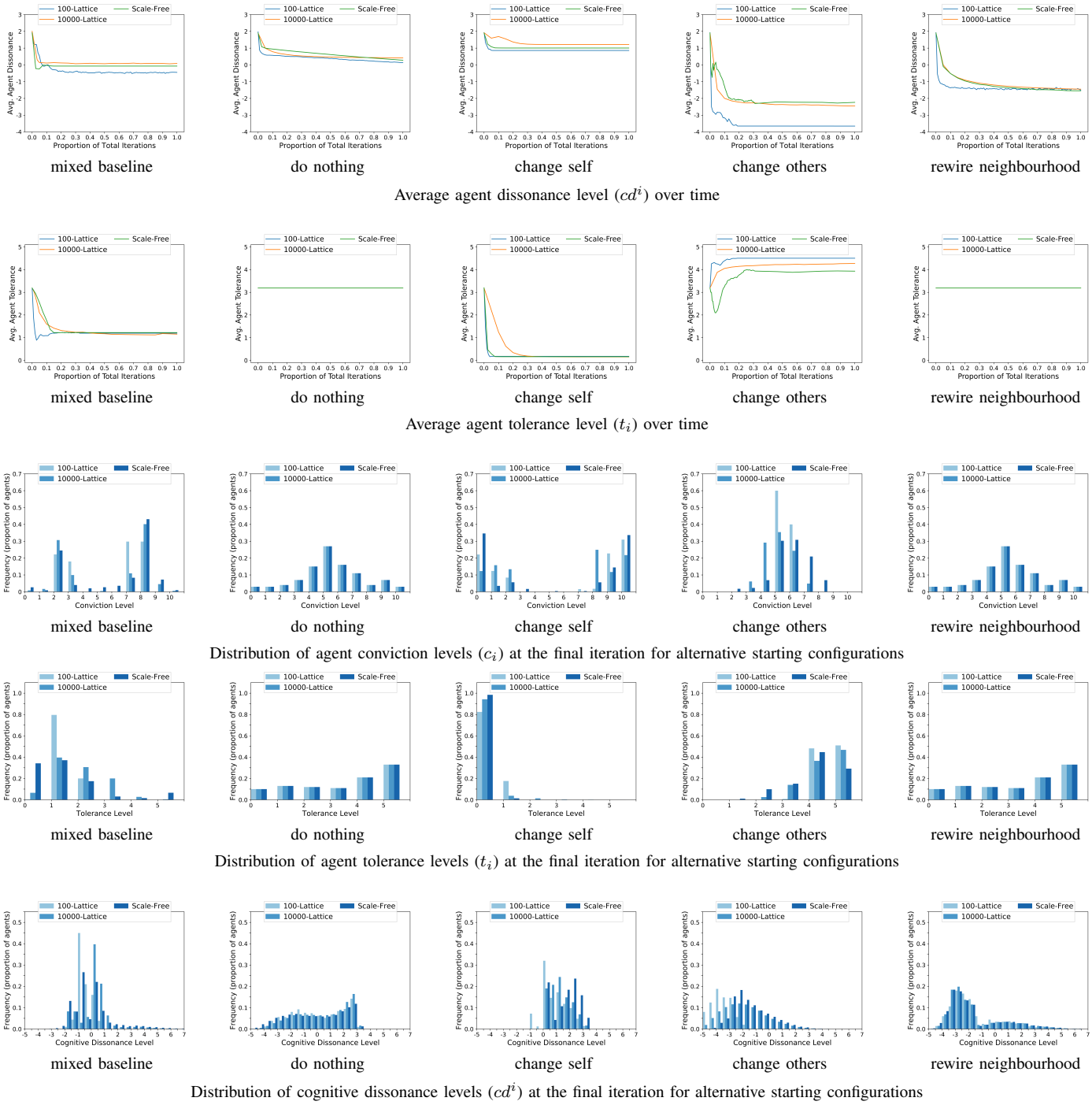
Index for conviction/tolerance levels by colour.

Fig. 2. An example of conviction and tolerance values at the end of a simulation, starting from a 100-agent lattice configuration across Scenarios A: mixed baseline, B: do nothing, C: change self, D: change others, E: rewire neighbourhood. Network nodes are labelled by conviction or tolerance level. To aid visualisation, colours are also used to represent each conviction and tolerance level.

mutation ($m = 0.001$). In practice it would be a special case in which human agents are unresponsive to cognitive dissonance, but modelling this scenario allows us to observe the network implications that result. Rewiring alone is insufficient to completely remove cognitive dissonance, which is due to links connecting agents with conviction outside of their tolerance levels. However the small amount of rewiring that is involved is sufficient to significantly reduce the overall cognitive dissonance (Table II) and disrupts the starting configurations, bringing all three types towards a random network (e.g., an Erdos-Renyi construction). While the diversity of alternative conviction levels is retained, alongside diversity of tolerance, the rewiring introduces shorter paths for the lattice starting configurations, while increasing shortest path length for the scale-free starting configuration. Similar frequency distributions of cognitive dissonance, conviction and tolerance (Figure 3) are evident from the alternative starting configurations.

Scenario C - Change Self

In this scenario the thresholds are configured to promote assimilation with others. For example, for the 10000-agent scale-free starting configuration, this results in 0.10% of actions being do nothing, 99.47% change self and 0.43% of actions are to rewire. This represents a conformist behavioural response that involves an agent taking an alternative conviction from their neighbourhood with a bias towards reducing their own cognitive dissonance. The results show that more extreme convictions take hold in the population as a consequence, with agents effectively using the stronger conviction levels to reduce the uncertainty. The results for the network are striking - strong opposing clusters are evident with the networks struggling to retain overall connectivity (see Figure 2). As agents become more polarised in their conviction levels, they



Average agent dissonance level (cd^i) over time

Average agent tolerance level (t_i) over time

Distribution of agent conviction levels (c_i) at the final iteration for alternative starting configurations

Distribution of agent tolerance levels (t_i) at the final iteration for alternative starting configurations

Distribution of cognitive dissonance levels (cd^i) at the final iteration for alternative starting configurations

Fig. 3. Average agent dissonance and tolerance levels over time, and final distributions of agent conviction, tolerance and cognitive dissonance from alternative starting configurations for Scenarios A, B, C, D, and E.

become less tolerant making it harder for cognitive dissonance to be removed. As such some cognitive dissonance remains in the network, principally through agents having neighbours with polarised alternative conviction levels.

In Figure 3, similar distributions of conviction and tolerance results are evident from all three starting configurations, showing the tendency towards opposing convictions and low tolerance. In these circumstances, an agent cannot resolve cognitive dissonance without rewiring, which only happens infrequently in this scenario. Figure 4 demonstrates how the 100-agent lattice progresses towards bridging between alternative conviction levels, with tolerance levels progressively reducing.

Scenario D - Change Other

The thresholds used to promote the "change other" behaviour focus on the change other action. For example, when starting with the 10,000-agent scale-free networks, 0.63% of actions involve do nothing, 97.33% change other and 2.04% rewire. Unlike the previous scenarios, this results in dynamics where agents are assumed effective in persuading others,

TABLE II
NETWORK METRICS FOR SCENARIOS A: MIXED BASELINE, B: DO NOTHING, C: CHANGE SELF, D: CHANGE OTHERS, E: REWIRE NEIGHBOURHOOD, AT THE END OF THE SIMULATION. L_{100} INDICATES THE 100-AGENT LATTICE, L_{10^4} INDICATES THE 10,000-AGENT LATTICE AND SF INDICATES THE 10,000 NODE SCALE-FREE NETWORK STARTING CONFIGURATIONS.

Measure	Starting Configuration			Scenario A: mixed baseline			Scenario B: do nothing			Scenario C: change self			Scenario D: Change others			Scenario E: rewire		
	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF
Average Cognitive Dissonance	1.926	1.914	1.891	-0.448	0.084	-0.072	0.141	0.413	0.278	0.858	1.215	1.005	-3.652	-2.453	-2.236	-1.502	-1.449	-1.546
Number of Alternative Conviction Levels	11.00	11.00	11.00	5.000	11.00	11.00	11.00	11.00	11.00	7.000	9.000	9.000	2.000	9.000	8.000	11.00	11.00	11.00
Average Conviction	5.250	5.250	5.250	5.454	5.409	5.562	5.250	5.250	5.250	5.808	5.656	5.178	5.400	4.925	5.772	5.250	5.250	5.250
Number of Alternative Tolerance Levels	6.000	6.000	6.000	3.000	6.000	6.00	6.000	6.00	6.00	2.000	5.000	4.000	3.000	5.000	5.000	6.000	6.000	6.000
Average Tolerance	3.190	3.190	3.190	1.196	1.152	1.223	3.190	3.190	3.190	0.176	0.159	0.152	4.502	4.272	3.931	3.190	3.190	3.190
Average Clustering	0.000	0.000	0.988	0.063	0.000	0.007	0.041	0.000	0.500	0.066	0.000	0.233	0.020	0.000	0.001	0.144	0.000	0.263
Assortativity by Conviction Levels	0.021	0.032	0.000	0.109	0.316	0.360	0.043	0.050	0.084	0.597	0.510	0.420	0.184	0.298	0.130	0.448	0.469	0.510
Average Local Diversity	2.446	2.403	2.303	0.319	0.570	0.643	1.619	1.758	1.471	0.697	0.745	0.796	0.340	0.543	0.551	0.731	0.708	0.668
Average Similarity	0.511	0.499	0.500	0.437	0.435	0.446	0.481	0.485	0.479	0.472	0.473	0.465	0.481	0.480	0.430	0.448	0.437	0.442
Average Shortest Path	6.667	66.67	1.242	3.671	7.610	6.989	3.834	8.250	11.41	4.220	7.682	10.24	4.234	8.402	6.677	3.953	7.204	8.842

mediated by similarity of fixed traits. In contrast to change self, change other provides a mechanism through which agents can become more tolerant. This takes hold through the dominant 'neutral' agents in the initial population, who are clustered on the mid-conviction level with high tolerance (see Figure 3). The result is a population that converges towards a uniform but neutral conviction level that has a high associated tolerance level. As a consequence cognitive dissonance is eliminated on aggregate (Table II) with similar distributions of cognitive dissonance achieved irrespective of the starting configuration (see Figure 3). Network connectivity exhibits a reduction in average shortest path length that is similar to Scenario B (do nothing) for the lattice starting configurations. Figure 4 shows how alternative conviction and tolerance levels rapidly diminish within the first few thousand iterations of response behaviour for the 100-agent lattice starting configuration.

Scenario E - Rewire

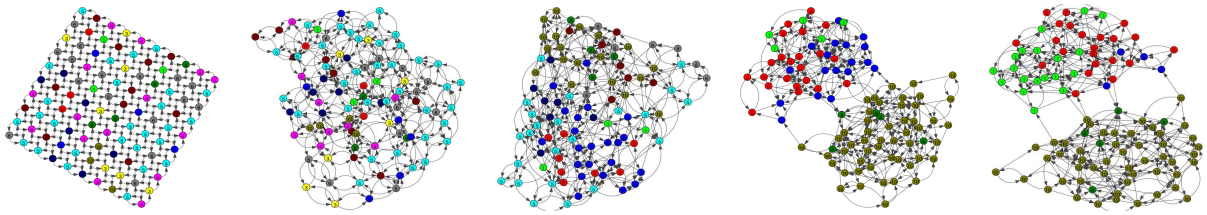
This scenario has thresholds designed to promote rewiring (for example 20.17% of actions are do nothing, 79.83% are rewire, starting from the 10,000-agent scale-free configuration). This behavioural response retains the diversity of conviction in the population, and focuses on restructuring the relationships to increase clustering around common and neighbouring conviction levels - see Figure 2 for an example. This typically occurs for conviction levels in the range 3-7 because these are more prevalent. This is effective in removing aggregate cognitive dissonance (Table II) with similar distributions of cognitive dissonance achieved for all starting configurations (Figure 3). Note that each agent retains the same conviction-tolerance level throughout. As a result of increased clustering for the lattice starting configurations, shortest path lengths are reduced, while rewiring disrupts the low path lengths for the scale-free starting configuration. Figure 4 demonstrates how rewiring progressively evolves the clusters of conviction and tolerance for the 100-agent lattice starting configuration, which clearly begin to establish themselves within the first 1,000 iterations. These clusters provide stability for agents, by reducing or eliminating cognitive dissonance.

TABLE III
SCENARIOS WHERE SPECIFIC SUB-POPULATIONS HAVE AMENDED RESPONSE BEHAVIOURS. THIS ASSUMES A MIXED BASELINE $T_0 = -3$, $T_1 = 0$, $T_2 = 3$ (SCENARIO A), UNLESS AMENDED BY THE IMPOSED RESPONSE BEHAVIOUR. NEUTRAL AGENTS ARE THOSE WITH CONVICTION LEVEL 5 AND TOLERANCE LEVEL 5. EXTREME AGENTS HAVE CONVICTION LEVEL OF 0 OR 10.

Scenario	Sub-population	Behaviour imposed
F	Neutral Agents	Do not change others, by using $T_0 = -3$, $T_1 = T_2 = 3$
G	Extreme Agents	Prevented from being changed by others
H	Neutral Agents	Do not change self and prevented from being changed by others (stubbornness)
I	Neutral and Extreme Agents	Prevented from being changed by others
J	Neutral and Extreme Agents	Neutral agents are stubborn. Extreme agents are prevented from being changed by others

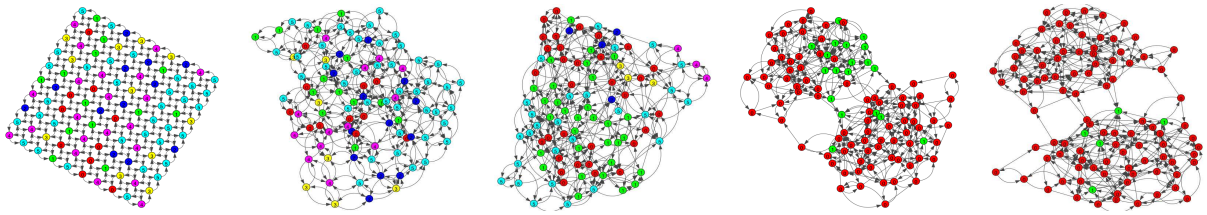
B. Heterogeneous Agent Responses to Cognitive Dissonance

In the previous Section all agents were assumed to act identically in response to cognitive dissonance. In this Section we relax this assumption, and consider that i) some agents may act with different thresholds T_0, T_1 and T_2 to trigger response behaviours; ii) individual differences will affect whether agents are persuaded by the 'change others' action. We explore these issues by assuming the wider population adopts the mixed baseline thresholds for response to cognitive dissonance (Scenario A, Table I), while experimenting with the thresholds for particular subsets of agents, or limiting their change behaviour. A summary of the scenarios that we consider in this Section is presented in Table III.



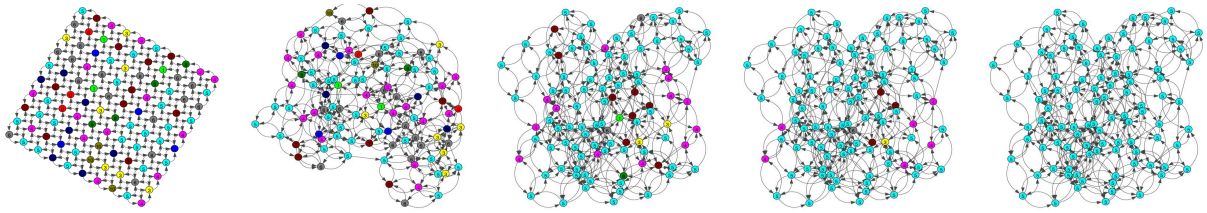
iteration 0 iteration 100 iteration 1000 iteration 3,000 iteration 10,000

Change self: Conviction



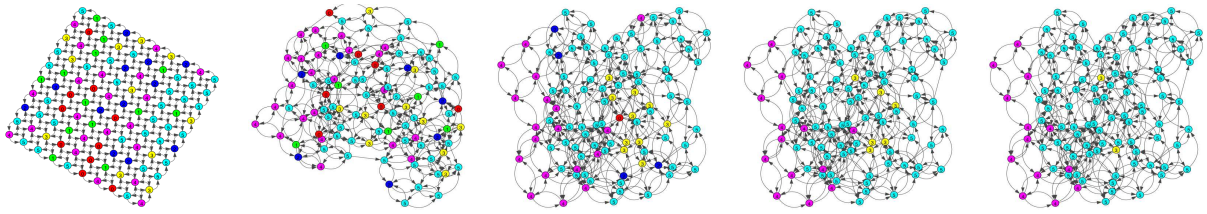
iteration 0 iteration 100 iteration 1,000 iteration 3,000 iteration 10,000

Change self: Tolerance



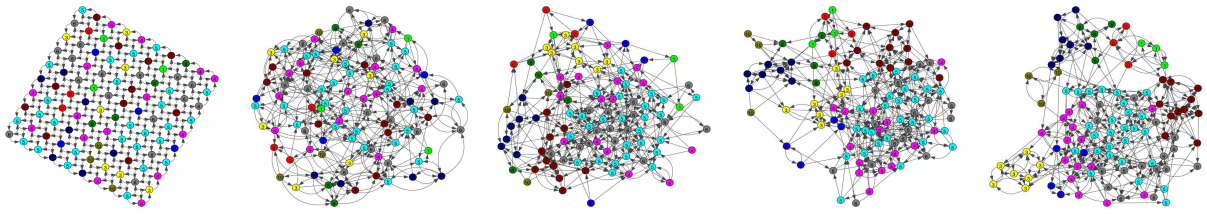
iteration 0 iteration 100 iteration 1,000 iteration 3,000 iteration 10,000

Change Others: Conviction



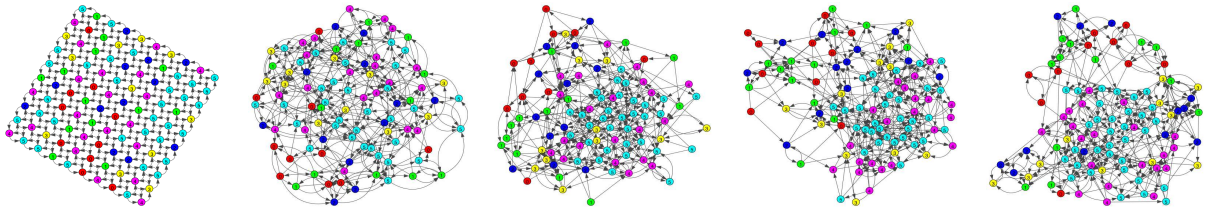
iteration 0 iteration 100 iteration 1,000 iteration 3,000 iteration 10,000

Change Others: Tolerance



iteration 0 iteration 100 iteration 1,000 iteration 3,000 iteration 10,000

Rewire Neighbourhood: Conviction



iteration 0 iteration 100 iteration 1,000 iteration 3,000 iteration 10,000

Rewire Neighbourhood: Tolerance

Fig. 4. Examples of the evolution of network structure from a 100-agent lattice, for conviction and tolerance over 10,000 iterations for the key response mechanisms - Scenario C: change-self, D: change others and E: rewire neighbourhood. Network nodes are labelled by conviction or tolerance level. To aid visualisation, colours are used to represent conviction and tolerance levels as defined in Figures 2 and 5.

Scenario F - The ability of neutral agents to persuade others mitigates polarisation towards extreme conviction levels across the population

The results from Scenario D have shown that the ability of neutral agents (i.e., mid-conviction - level 5, high tolerance - level 5) to cascade their attributes through the 'change others' response to cognitive dissonance. The comparison with the other scenarios in Section IV-A also shows that this is the only mechanism through which higher tolerance levels can be propagated across the population, and is driven by the presence of neutral agents.

Consequently we experiment with removing the 'change other' option from neutral agents. While assuming the population adopts the mixed baseline ($T_0 = -3, T_1 = 0, T_2 = 3$), we restrict the neutral agents to not being able to persuade others ($T_0 = -3, T_1 = T_2 = 3$). This means that neutral agents only change their conviction and tolerance levels through the 'change self' mechanism, or from persuasion by the most extreme agents. The results show there is greater aggregate cognitive dissonance (Table IV) as compared to Scenario A, reflective of the reduced role of neutral agents. Figure 5 shows a representative network example for Scenario F. We note that as compared to the mixed baseline (Figure 3), Scenario F maintains low tolerance levels (Figure 6). Very similar results are evident from the alternative starting configurations.

Scenario G - When extreme agents cannot be persuaded to change conviction, the network is susceptible to extreme polarisation

In this experiment, while assuming the population adopts the mixed baseline ($T_0 = -3, T_1 = 0, T_2 = 3$), the extreme conviction agents (with conviction 0 or 10) are prevented from being changed by others. This is consistent with extreme conviction and low tolerance, and restricts the potential for assimilation of such members. Interestingly the behaviour means that extreme conviction agents can only assimilate with themselves, while other agents in the population progressively get drawn towards the extreme conviction levels, mainly from assimilation to reduce dissonance. For example, when starting with the 10,000-agent scale-free networks, the behavioural responses are: 'do nothing': 1.34%; 'change self': 86.99%; 'change others': 10.65% and 'rewire': 1.02%. This leads to a strongly polarised situation, where extreme conviction agents cluster at alternative ends of the spectrum (see Figure 5 for an example). This occurs across all three types of starting configuration, culminating in similar results concerning aggregate cognitive dissonance (Table IV) and similar distributions of conviction and tolerance (see Figure 6).

Scenario H - When neutral agents don't assimilate and can't be persuaded by others, then the population maintains greater diversity of conviction levels

In Section IV-A, the propagation of the neutral agent was shown to be strongly aligned with the 'change others' response to cognitive dissonance. Here we determine the effects of stopping the changes to neutral agents (i.e., mid-conviction - level 5, high tolerance - level 5), both through assimilating towards others and being persuaded by others, while allowing them to act through the mechanism of persuasion only. This aligns with such agents having a strong disposition to retain neutral characteristics and an example result is given in Figure 5. In contrast to the effects of change others (Figure 2 and Figure 3), preventing neutral agents from assimilation allows them to persist and this enables opposing minority conviction levels to establish themselves in the network (see Figure 6). Interestingly, aggregate cognitive dissonance is eliminated in this scenario (Table IV), with the stability of the neutral agents providing an effect that allows other conviction levels to establish and connect themselves indirectly to the wider population. Outside of the fixed neutral agents, the mechanisms at play involve 'change others' taking most hold. For example, starting from the 10,000-agent scale-free scenario, we see 'do nothing': 1.0%, 'change self': 37:38%, 'change others': 59.50%, and 'rewire': 2.12%.

Scenario I - When neutral and extreme agents can't be persuaded, neutral agents bridge the network, but cognitive dissonance remains

The conclusions in the previous experiment (Section IV-B) assume that extreme agents are still susceptible to persuasion. When this is revoked then flexibility is reduced, as shown by the example in Figure 5 for Scenario I. There is a tendency for those agents that are open to assimilation to migrate to extreme conviction levels with 'change self' being a dominant response to cognitive dissonance (for example, starting from the 10,000-agent scale-free starting configuration, 'do nothing': 2.19%, 'change self': 65.07%, 'change others': 30.96%, 'rewire': 1.78%). Meanwhile the neutral agents remain in a bridging role that maintains overall connectivity. This arrangement is evident in conviction and tolerance levels (see Figure 6). The low tolerance to which agents migrate leads to cognitive dissonance remaining at a substantial level (Table IV).

Scenario J - When both neutral and extreme agents can't be persuaded and neutral agents also do not assimilate, clusters of polarisation embed in a neutral population

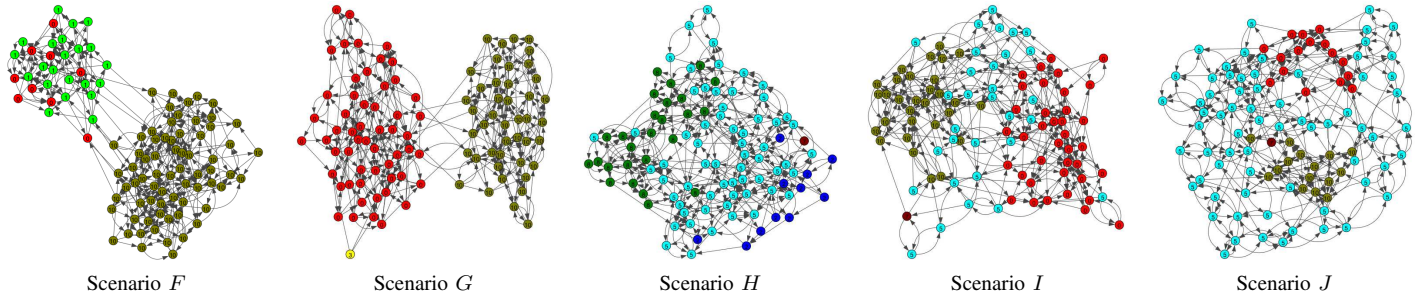
In this scenario, the previous conditions (Section IV-B) are imposed, but with the additional constraint that neutral agents do not perform 'change self' in response to cognitive dissonance. This presents the neutral agents with a possibility of changing others who are of a moderate but not extreme conviction level. In this context neutral agents are successful in their persuasion of such agents and a distinctive tripartite network of conviction arises, but with neutral agents dominating. Clusters of agents having opposing extreme conviction become embedded in an extensive number of neutral agents and this arrangement allows cognitive dissonance to be removed, on aggregate, from the network with the neutral agents facilitating connectivity to the respective extreme groups (Figure 5). This enables low aggregate cognitive dissonance levels (Table IV) and patterns of conviction and tolerance (see Figure 6) that are similar, independent of the initial starting configuration.

C. Meta-Analysis Across Existing Data

A variety of experiments and studies from the wider literature focus on observing the consequences of cognitive dissonance through data, as generated at a group or network level. In Table V, we present an analysis of related papers that observe social interactions in the context of discord or

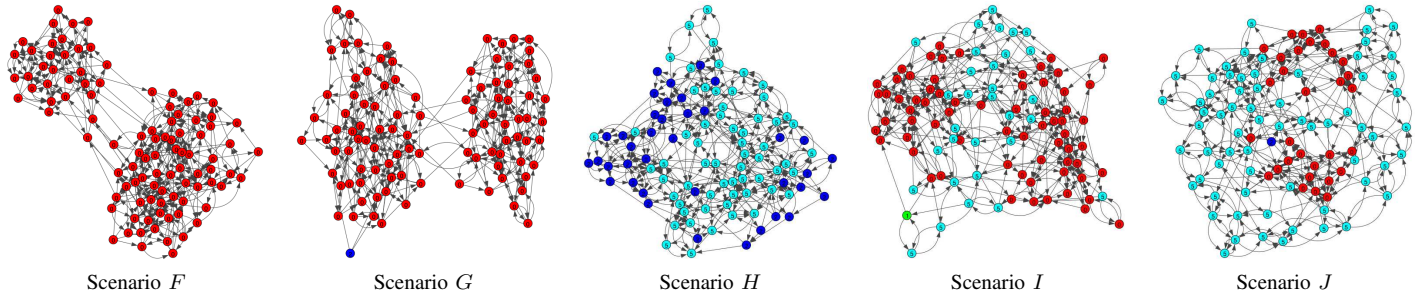
TABLE IV
NETWORK METRICS FOR SCENARIOS F, G, H, I AND J AT THE END OF THE SIMULATION. L_{100} INDICATES THE 100-AGENT LATTICE, L_{10^4} INDICATES THE 10,000-AGENT LATTICE AND SF INDICATES THE 10,000 NODE SCALE-FREE NETWORK STARTING CONFIGURATIONS.

Measure	Starting Configuration			Scenario F			Scenario G			Scenario H			Scenario I			Scenario J		
	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF	L_{100}	L_{10^4}	SF
Average Cognitive Dissonance	1.926	1.914	1.891	0.551	0.593	0.537	0.202	0.184	0.198	-1.463	-1.394	-1.552	0.927	0.898	0.966	-0.907	-0.764	-0.824
Number of Alternative Conviction Levels	11.00	11.00	11.00	6.000	11.00	8.000	4.000	11.00	11.00	7.000	11.000	11.00	7.000	11.00	11.00	7.000	11.00	11.00
Average Conviction	5.250	5.250	5.250	5.854	5.842	5.433	5.689	5.574	4.753	5.528	5.137	5.738	5.622	5.694	5.434	5.146	5.264	5.356
Number of Alternative Tolerance Levels	6.000	6.000	6.000	1.000	5.000	5.000	4.000	6.000	6.000	5.000	6.000	6.000	5.000	6.000	6.000	6.000	6.000	6.000
Average Tolerance	3.190	3.190	3.190	0.000	0.078	0.000	0.018	0.049	0.038	3.384	3.351	3.241	1.725	1.556	1.787	3.306	3.361	3.395
Average Clustering	0.000	0.000	0.988	0.074	0.000	0.001	0.000	0.000	0.002	0.055	0.000	0.355	0.085	0.000	0.141	0.073	0.000	0.292
Assortativity by Conviction Levels	0.021	0.032	0.000	0.718	0.614	0.522	0.920	0.762	0.932	0.313	0.377	0.480	0.533	0.546	0.688	0.572	0.473	0.580
Average Local Diversity	2.446	2.403	2.303	0.554	0.572	0.485	0.268	0.396	0.226	0.882	0.710	0.707	1.469	1.426	1.185	1.100	0.937	0.841
Average Similarity	0.511	0.499	0.500	0.451	0.443	0.450	0.449	0.425	0.443	0.439	0.454	0.462	0.469	0.458	0.477	0.484	0.456	0.478
Average Shortest Path	6.667	66.67	1.242	3.906	7.013	6.701	4.244	7.371	7.847	3.836	7.774	9.122	3.961	7.792	8.453	4.100	7.794	9.29



Scenario F Scenario G Scenario H Scenario I Scenario J

Conviction levels held by agents at the final iteration, after starting from a 100-agent lattice configuration.



Scenario F Scenario G Scenario H Scenario I Scenario J

Tolerance levels held by agents at the final iteration, after starting from a 100-agent lattice configuration.

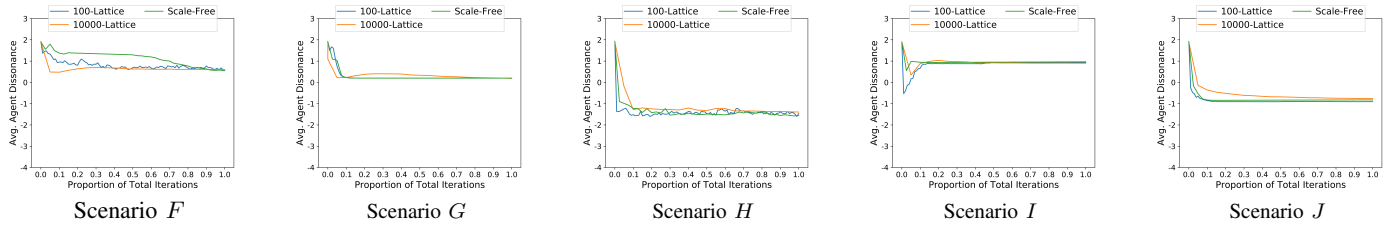
Index for conviction/tolerance levels by colour.

Fig. 5. An example of conviction and tolerance values at the end of a simulation, starting from a 100-agent lattice configuration, across Scenarios F, G, H, I, and J. Network nodes are labelled by conviction or tolerance level. To aid visualisation, colours are used to represent each conviction and tolerance level.

contention, and the characteristics of networks that result. In particular, we examine how different contributions relate to behavioural responses concerning cognitive dissonance, as presented through our model.

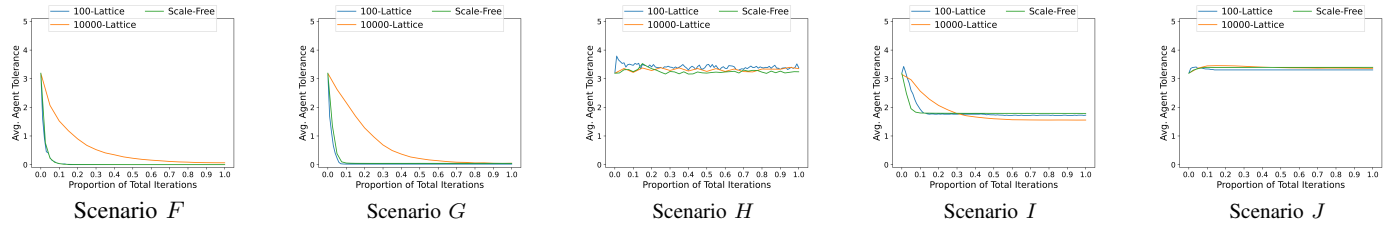
Broadly speaking, academic contributions in this area occur in three forms: firstly through controlled experiments, where exogenous variables are managed, alongside controlled participation. Secondly, contributions characterise real-world social media, where data from online interactions allow behavioural responses to cognitive dissonance to be observed. In these circumstances, because interactions are occurring "in-the-wild", exogenous influences cannot be controlled, however the resulting behavioural effects are often observable at large scale. Finally agent-based models, based on simulated agent interactions, effectively play out a sequence of responses and counter responses based on initial starting configurations, which may be random or emulated from data.

In terms of explicit controlled participation, numerous studies invoke experimental controls [57], [68], [59], [28], [60], [86], [64], [49], [63], [13]. These have mainly focussed on



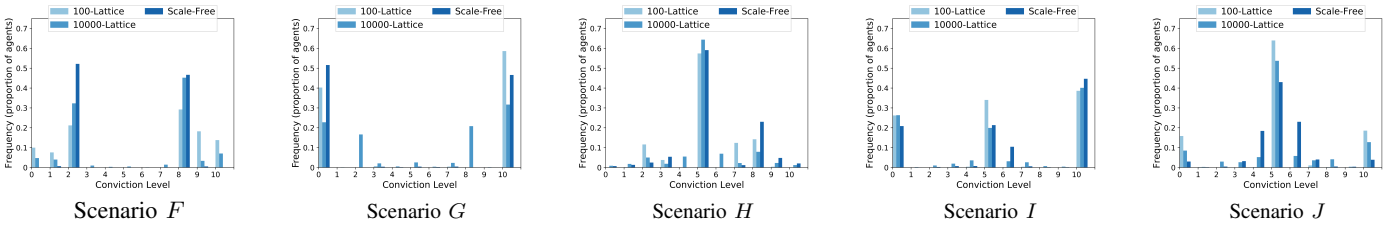
Scenario F Scenario G Scenario H Scenario I Scenario J

Average agent dissonance level (cd^i) over time



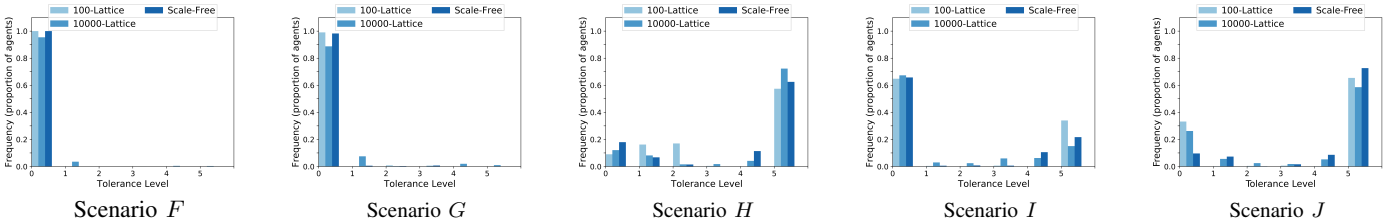
Scenario F Scenario G Scenario H Scenario I Scenario J

Average agent tolerance level (t_i) over time



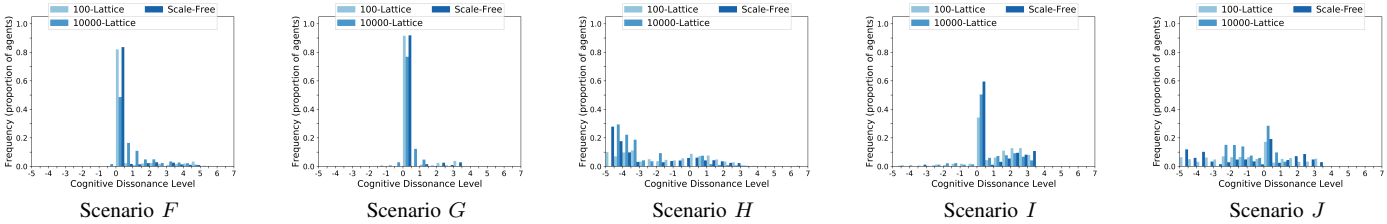
Scenario F Scenario G Scenario H Scenario I Scenario J

Distribution of agent conviction levels (c_i) at the final iteration for alternative starting configurations



Scenario F Scenario G Scenario H Scenario I Scenario J

Distribution of agent tolerance levels (t_i) at the final iteration for alternative starting configurations



Scenario F Scenario G Scenario H Scenario I Scenario J

Distribution of agent cognitive dissonance levels (cd^i) at the final iteration for alternative starting configurations

Fig. 6. Average agent dissonance, tolerance and cognitive dissonance levels over time, and final distributions of agent conviction and tolerance from alternative starting configurations for Scenarios F, G, H, I, and J.

lab-based experimentation, involving manipulation of the conditions that influence individual behaviour, such as providing additional information (e.g., [13]), or triggers to test opinion swapping [64], [49], [63] based on social influence [77], [54].

Beyond this, engagement with large-scale social media normally takes two forms, either fitting models to observed social media data (e.g., [22], [44], [52], [78], [71]), or in-depth observation of social media in its own right, which frequently characterises the extent and depth of polarisation (e.g., [67], [6], [88], [35], [30], [33], [45], [2], [18], [50], [39], [8]). In terms of model fitting, this is directed at prediction of opinion formation, and data from Twitter has frequently been used. While not explicitly assessing the extent of cognitive dissonance, many models incorporate a probability of opinion copying, based on influence from others. This typically involves a static network representation (i.e., links remain fixed overtime) and the pursuit of circumstances where polarisation becomes invoked, with [22] and [78] being particularly interesting examples.

Alternatively, social media is frequently observed and assessed to determine its characteristics with respect to the underlying community's polarisation. This includes structural

considerations (such as the extent of clustering) as well as content analysis and behavioural indicators (e.g., retweets, online friends or survey data). In combination, these factors can contribute to echo-chambers (e.g., [6], [33]) that amplify polarisation, as reinforced by selective exposure [43], [70], [4], [31]. However, this is not always the case [84], [12], [2], and may depend on the type of social media data [18], [87].

With respect to agent-based models, the utility of this approach concerns observing how isolated behaviours, or combinations of behaviours interact with each other to create collective effects. This can assist in understanding hypothetical scenarios or in answering specific questions without the presence of exogenous factors. Such investigation requires initialisation of agent configurations, so that alternative agent actions and reactions can be observed in context. Initial agent configurations typically involve generating data so that parameter sensitivity can be assessed. This classification captures the contribution of this paper, and is a widely used in the literature, as evidenced by [21], [20], [51], [82], [52], [22], [44], [78], [71].

The analysis provided in Table V highlights that the inclusion of cognitive dissonance, either implicitly or explicitly, is highly prevalent in scenarios addressing social networks and contention. When this isn't the case, the related work in Table V focusses on the characterisation of polarisation. A wide range of phenomena are considered throughout this related work, which we categorise to include assessments of in-out group formation, in-group diversity, in-group uniformity, selective exposure, extent of dissonance, mechanisms of social influence, change to social structure, social contagion and opinion switching.

Inclusion of response mechanisms to cognitive dissonance are also summarised in Table V (assimilation, persuasion, do nothing, rewiring one's social network, stubbornness - cannot be persuaded). These are a retrospective mapping of external works to the response mechanisms studied in this paper, and in some cases these are open to interpretation. Interestingly, assimilation, where individuals adjust their own views to reduce inconsistency with others, dominants as a primary response mechanism for dissonance across all the main types of study, and this is incorporated both implicitly and explicitly. Much less prevalent is the explicit inclusion of persuasion [82] - this reflects the challenge of disaggregating persuasion from assimilation based on observation of social media, without further participatory input. Also, rewiring is mostly observed in participatory lab-based studies, and this has been identified as a gap in the literature on opinion dynamics [51].

Finally in Table V, we present an assessment of the correspondence between related work and the most relevant of our scenarios A to J. This indicates the scenarios that are likely to provide insights into the underlying and related mechanisms observed in the specific literature. While this represents a subjective assessment, it provides value in identifying the phenomena that result when the mechanisms function in isolation from exogenous factors. This gives an independent point of reference on specific effects caused by behavioural responses to cognitive dissonance.

V. Discussion and Conclusions

The results give insight into how vicarious dissonance manifests itself, through individual responses to cognitive dissonance that are felt through the social network. We note that our model is intentionally simple, allowing for the observation of collective effects in the absence of external factors. Even with this simplicity, there are still thousands of different parameters settings that can be assessed, leading to large numbers of potential scenarios. In light of this, the experiments that have been considered focus on identifying the key mechanisms at play, and their effects. Fundamental to this is the initial distribution of cognitive dissonance across the population of agents. This is caused by both the initial distributions of conviction (and tolerance) across agents, and how these are interconnected through the network. Our assumptions throughout are that more extreme conviction levels are tied with lower tolerance.

The alternative responses to cognitive dissonance result in distinct effects. Preservation of alternative conviction levels only occur through rewiring behaviour, which is perhaps counter-intuitive given that it is an extreme response. However this behaviour accelerates clustering around more extreme conviction levels, supporting polarisation. In other words, rewiring reduces cognitive dissonance through clustering with like-minded others. This is seen in Scenario E (Section IV-A where the networks become heavily disrupted with short paths existing within particular conviction levels). The other responses to cognitive dissonance represent self-assimilation (change self) or persuasion (change others). Assimilation works in favour of those with more extreme conviction levels, and is invoked when the cognitive dissonance experienced is relatively low (i.e., before change others or rewiring). All other things being equal, this is more likely to be the case for neutral agents where tolerance is high. As a consequence, self-change results in a diminishing number of neutral agents that migrate towards more extreme conviction levels where tolerance is low. This is seen in Scenario C (Section IV-A) and is consistent with reducing the number of conviction levels in the population. In contrast, the persuasion behaviour (i.e., change others) works with the opposite effect, being a mechanism through which higher levels of tolerance can be spread across a population. However this is dependent on the target agent, whose change is sought, being open to that change. When our experiments (Section IV-A, scenario D) assume this is the case, and when this is the main response to cognitive dissonance, we readily see convergence to higher tolerance and neutral conviction levels. This is because agents with more extreme conviction levels (and therefore lower tolerance levels) tend not to sustain many links with neutral (high tolerance) agents, while neutral agents can sustain links to extreme agents with less cognitive dissonance. Consequently neutral agents are more likely to use this mechanism, which allows neutral conviction to spread. This occurs through prioritisation with those that are more similar in terms of traits, and then cascades through the population.

With two of the three response behaviours to cognitive dissonance (i.e., assimilate and rewire) tending to promote

TABLE V

COMPARISON OF KEY FEATURES OF PREVIOUS STUDIES. LB=LAB-BASED STUDY, OS=ONLINE SURVEY, SM= ONLINE SOCIAL MEDIA DATA, AB=AGENT-BASED SIMULATION. IMP.=IMPLICIT, EXP.=EXPLICIT. SO=SINGLE OBSERVATIONS, MO=MULTIPLE OBSERVATIONS, AO=AGGREGATED OBSERVATIONS. IF=IN-OUT GROUP FORMATION, ID=IN-GROUP DIVERSITY, IU=IN-GROUP UNIFORMITY, SE=SELECTIVE EXPOSURE, ED=EXTENT OF DISSONANCE, SI=MECHANISMS OF SOCIAL INFLUENCE, SS=CHANGE TO SOCIAL STRUCTURE, SC=SOCIAL CONTAGION, OP=OPINION SWITCHING. AS=ASSIMILATION, PS=PERSUASION, DN=DO NOTHING, RW=REWIRING, ST=STUBBORNESS (CANNOT BE PERSUADED). N/A=NOT APPLICABLE.

	Components				**Primary Topic/Context**	**Cog. Diss.**	**Toler-ance**	**Polari-sation**	**Data Temporality**			**Key Contributions**									**Response to Dissonance**					**Relevant Scenarios**
	LB	OS	SM	AB					SO	MO	AO	IF	ID	IU	SE	ED	SI	SS	SC	OP	AS	PS	DN	RW	ST	
[64]	•				in-lab dissonance	exp.	N/A	N/A		•						•				•	•		•			*B,C*
[49]	•				in-lab dissonance	exp.	N/A	N/A		•						•				•	•		•			*B,C*
[63]	•				in-lab dissonance	exp.	N/A	N/A		•						•				•	•		•			*B,C*
[57]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•	•	•	•		*B,C,D,E*
[68]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•		•	•		*B,C,E*
[59]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•		•	•		*B,C,E*
[28]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•		•	•		*B,C,E*
[60]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•		•	•		*B,C,E*
[86]	•				in-lab dissonance	exp.	imp.	imp.		•							•	•		•	•		•	•		*B,C,E*
[73]	•				moral issues	exp.	imp.	imp.	•				•	•							•		•			*B,C*
[23]	•				other politics	exp.	imp.	imp.		•		•			•		•		•			•		•		*C,E*
[80]	•				other politics	exp.	exp.	imp.		•		•					•						•	•		*B,E*
[79]	•				other politics	exp.	exp.	imp.		•		•					•						•	•		*B,E*
[29]	•				moral issues	exp.	exp.	imp.		•		•					•						•	•		*B,E*
[13]	•				other politics	imp.	imp.	exp.		•							•	•		•	•	•	•	•		*B,C,E*
[72]		•			moral issues	exp.	imp.	exp.	•				•	•		•		•			•		•	•		*B,C,E*
[70]		•			other politics	exp.	N/A	N/A	•						•	•										*N/A*
[81]		•			other politics	exp.	N/A	N/A		•					•	•										*N/A*
[4]		•			other politics	exp.	imp.	imp.		•	•	•		•	•	•		•			•					*C*
[31]		•			elections	imp.	imp.	exp.		•		•		•				•			•			•		*C,E*
[50]		•			other politics	imp.	imp.	exp.	•			•	•										•			*B*
[77]		•			consumerism	exp.	imp.	N/A		•						•					•		•			*B,C*
[54]		•			consumerism	exp.	imp.	N/A		•						•				•	•		•			*B,C*
[43]		•			emotive content	exp.	N/A	N/A	•						•	•										*N/A*
[45]		•			elections	imp.	imp.	exp.	•			•		•						•	•	•			•	*C,F,G,I,J*
[62]		•			other politics	exp.	imp.	imp.	•							•					•		•			*B,C*
[36]		•			other politics	exp.	imp.	imp.	•						•	•					•		•			*B,C*
[53]		•			moral issues	exp.	N/A	N/A	•							•										*N/A*
[8]		•			other politics	imp.	imp.	exp.			•	•	•										•			*B*
[6]		•	•		other politics	exp.	imp.	imp.	•			•		•							•					*C*
[52]		•	•	•	other politics	imp.	exp.	exp.	•	•		•					•	•			•	•		•		*A,C,D,E*
[39]			•		other politics	imp.	imp.	exp.			•	•		•							•					*C*
[67]			•		other politics	imp.	imp.	exp.		•		•		•							•					*C*
[14]			•		extremism	N/A	N/A	imp.		•	•								•		•	•				*A*
[85]			•		extremism	N/A	N/A	imp.		•	•								•		•	•				*A*
[88]			•		moral issues	imp.	imp.	exp.		•	•	•		•				•			•			•		*C,E*
[35]			•		elections	imp.	imp.	exp.			•	•		•						•	•					*C*
[34]			•		elections	imp.	imp.	exp.		•	•	•		•				•			•			•		*C,E*
[30]			•		elections	imp.	imp.	exp.		•	•	•		•				•			•			•		*C,E*
[33]			•		elections	imp.	imp.	exp.			•	•		•						•	•					*C*
[87]			•		other politics	imp.	imp.	exp.		•	•	•	•	•				•	•		•			•	•	*B,C,E*
[11]			•		moral issues	N/A	N/A	imp.		•	•									•	•	•				*A*
[10]			•		moral issues	N/A	N/A	imp.		•	•									•	•	•				*A*
[9]			•		moral issues	N/A	N/A	imp.		•	•									•	•	•				*A*
[2]			•		extremism	imp.	imp.	exp.	•			•	•	•							•			•		*B,C*
[18]			•		elections	imp.	imp.	exp.			•	•	•	•						•	•			•		*B,C*
[22]			•	•	moral issues	imp.	exp.	exp.		•		•		•			•				•	•				*A,C,D*
[44]			•	•	consumerism	imp.	imp.	imp.		•	•						•				•	•				*A,C,D*
[78]			•	•	consumerism	imp.	exp.	exp.		•	•	•					•	•			•	•		•	•	*C,E,F,G,I,J*
[71]			•	•	other	imp.	imp.	exp.		•		•		•			•	•			•	•			•	*C,F,G,I,J*
[21]				•	theoretical	imp.	exp.	exp.		•							•				•	•				*A,C,D*
[20]				•	theoretical	imp.	exp.	exp.		•							•				•	•				*A,C,D*
[51]				•	theoretical	exp.	exp.	exp.		•							•				•	•		•		*A,C,D,E*
[82]				•	theoretical	imp.	exp.	exp.		•	•						•				•	•				*A,C,D*

clustering and more extreme levels of conviction, it is evidence from the experiments that sustaining pluralism or the presence of neutrality is a considerable challenge. This is demonstrated for example by the mixed baseline, which convergences towards a limited range of conviction levels (e.g., Scenario A, Section IV-A). We note that this is critically governed by the points at which behavioural thresholds (T_0, T_1, T_2) are invoked. In reality these are governed by many external factors, from history, culture, government and social pressure, as well as individual dispositions.

Alongside the sensitivity to thresholds, the disposition of agents to be changed by others significantly influences the structure of the network and profile of the conviction within it. When sub populations become fixed with respect to their con-

viction levels, rewiring is invoked as a means to reduce cognitive dissonance, alongside assimilation, opening the potential for polarisation. This is seen in multiple scenarios, particularly when extreme conviction agents are involved (Scenarios F and G). Only when neutral agents remain fixed (Scenario H) do we see conditions where diversity in conviction levels being maintained. This reinforces the role and importance of neutral agents in achieving pluralism - such agents have the ability to reach diverse alternative convictions without invoking cognitive dissonance. Scenarios I and J further demonstrate this, showing how they can maintain connectivity through bridging (e.g., Scenario I) or through embedding (e.g., Scenario J) of more extreme sub-groups.

As configured (i.e., with complete randomisation at initialisation) we note that the fixed traits carried by the agents have a limited role in influencing how network structure emerges aligned to mitigating cognitive dissonance (average trait similarity in Tables II and IV). However this allows us to understand the dynamics associated with responses to cognitive dissonance in the absence of extraneous and confounding variables.

Finally, it is important to note that the initial network structure and scale of network have limited influence on the characteristics of the terminal configuration, provided the simulations are run for sufficiently long. The behaviours invoked by responses to cognitive dissonance are sufficient to dissolve structural features and reconfigure the initial networks with common characteristics. Despite starting from small lattices (100 agents), large lattices (10,000 agents) or scale-free networks (10,000 agents), similar results emerge, as reflected in the distributions of cognitive dissonance, tolerance and conviction. This phenomenon occurs because the response mechanisms to cognitive dissonance have some freedom to rewire the network and change conviction and tolerance levels, dependent on the particular response mechanism. This drives the initial network towards a stable configuration where cognitive dissonance is reduced to a tolerable level, as far as possible. Our results show that this occurs in different ways for different response behaviours. This observation is consistent with instances of polarisation and intolerance observed in different adversarial real-world scenarios, which for example, can often exhibit polarisation independent of the original social network structure.

A. Conclusions

We have shown that individual responses to cognitive dissonance aggregate to form interesting collective effects from a limited set of simple behaviours. Interestingly, the initial network configurations appear to hold little influence on the network characteristics that emerge. It is the behavioural responses of the agents in the network that invoke changes and drive the network towards specific features as a consequence of dissonance reduction. Particularly striking is the sensitivity to thresholds in terms of the different networks and conviction profiles that result, alongside the difficulties in maintaining diversity of conviction levels without polarisation and clustering. These issues resonate with characteristics seen in the human world, where phenomena such as populism are easily triggered while moderation and pluralism are less well disposed to cascade and are easily disrupted. The model introduced has been shown effective in bridging individual behaviour as a result of cognitive dissonance with with group-level behaviour. As far as we can establish, it is the first interdisciplinary contribution of this nature. We note that the model is highly configurable and extensible, allowing it to be used for particular scenarios or under bespoke assumptions for a particular situation, including dynamics relevant to an organisation or coalition.

Acknowledgement

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon. This research was also supported by the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government. L.D.T., G.B.C. and R.M.W. undertook this work as members of the Cardiff University Crime and Security Research Institute.

REFERENCES

[1] Victor Amelkin, Francesco Bullo, and Ambuj K Singh. Polar opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(11):5650–5665, 2017.

[2] Jisun An, Haewoon Kwak, Yelena Mejova, De Oger, Sonia Alonso Saenz, and Braulio Gomez Fortes. Are you charlie or ahmed? cultural pluralism in charlie hebdo response on twitter. *arXiv preprint arXiv:1603.00646*, 2016.

[3] Tibor Antal, Paul L Krapivsky, and Sidney Redner. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena*, 224(1-2):130–136, 2006.

[4] Natalia Aruguete and Ernesto Calvo. Time to# protest: Selective exposure, cascading activation, and framing in social media. *Journal of communication*, 68(3):480–502, 2018.

[5] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.

[6] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018.

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[8] Pablo Barberá. How social media reduces mass political polarization. evidence from germany, spain, and the us. *Job Market Paper, New York University*, 46, 2014.

[9] William J Brady, MJ Crockett, and Jay J Van Bavel. The mad model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 15(4):978–1010, 2020.

[10] William J Brady, Jay J Van Bavel, John Jost, and Julian Wills. An ideological asymmetry in the diffusion of moralized content among political elites. 2018.

[11] William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318, 2017.

[12] Jonathan Bright, Nahema Marchal, Bharath Ganesh, and Stevan Rudinac. Echo chambers exist!(but they're full of opposing views). *arXiv preprint arXiv:2001.11461*, 2020.

[13] Elena Buliga and Cara MacInnis. "how do you like them now?" expected reactions upon discovering that a friend is a political out-group member. *Journal of Social and Personal Relationships*, 37(10-11):2779–2801, 2020.

[14] Pete Burnap and Matthew L Williams. Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, 7(2):223–242, 2015.

[15] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[16] Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011.

[17] Damon Centola, Juan Carlos Gonzalez-Avella, Victor M Eguiluz, and Maxi San Miguel. Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6):905–929, 2007.

[18] Michael D Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. *Icwsm*, 133(26):89–96, 2011.

[19] William D Crano and Viviane Seyranian. How minorities prevail: The context/comparison–leniency contract model. *Journal of Social Issues*, 65(2):335–363, 2009.

[20] Rajkumar Das, Joarder Kamruzzaman, and Gour Karmakar. Consistency driven opinion formation modelling in presence of external sources. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2015.

[21] Rajkumar Das, Joarder Kamruzzaman, and Gour Karmakar. Who are convincing? an experience based opinion formation dynamics in online social networks. In *European Simulation and Modelling Conference 2016*, pages 167–173. EUROSIS-ETI, 2016.

[22] Rajkumar Das, Joarder Kamruzzaman, and Gour Karmakar. Opinion formation in online social networks: Exploiting predisposition, interaction, and credibility. *IEEE Transactions on Computational Social Systems*, 6(3):554–566, 2019.

[23] Scott Eidelman, Paul J Silvia, and Monica Biernat. Responding to deviance: Target exclusion and differential devaluation. *Personality and Social Psychology Bulletin*, 32(9):1153–1164, 2006.

[24] Andrew J Elliot and Patricia G Devine. On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of personality and social psychology*, 67(3):382, 1994.

[25] L Festinger. A theory of cognitive dissonance. stanford, calif.: Stanford university press. 1957.

[26] Andreas Flache. Between monoculture and cultural polarization: Agent-based models of the interplay of social influence and cultural diversity. *Journal of Archaeological Method and Theory*, 25(4):996–1023, 2018.

[27] Andreas Flache and Michael W Macy. Local convergence and global diversity: From interpersonal to social influence. *Journal of Conflict Resolution*, 55(6):970–995, 2011.

[28] Elizabeth S Focella, Jeff Stone, Nicholas C Fernandez, Joel Cooper, and Michael A Hogg. Vicarious hypocrisy: Bolstering attitudes and taking action after exposure to a hypocritical ingroup member. *Journal of Experimental Social Psychology*, 62:89–102, 2016.

[29] Immo Fritsche, Thomas Kessler, Amélie Mummendey, and Jörg Neumann. Minimal and maximal goal orientation and reactions to norm violations. *European Journal of Social Psychology*, 39(1):3–21, 2009.

[30] Kiran Garimella and Ingmar Weber. A long-term analysis of polarization on twitter. *arXiv preprint arXiv:1703.02769*, 2017.

[31] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. Me, my echo chamber, and i: introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference*, pages 823–831, 2018.

[32] Patrick Groeber, Jan Lorenz, and Frank Schweitzer. Dissonance minimization as a microfoundation of social influence in models of opinion formation. *The Journal of Mathematical Sociology*, 38(3):147–174, 2014.

[33] Anatoliy Gruzd and Jeffrey Roy. Investigating political polarization on twitter: A canadian perspective. *Policy & internet*, 6(1):28–45, 2014.

[34] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. A measure of polarization on social media networks based on community boundaries. In *ICWSM*, 2013.

[35] Anna Guimaraes, Liqiang Wang, and Gerhard Weikum. Us and them: Adversarial politics on twitter. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 872–877. IEEE, 2017.

[36] Rebecca A Hayes, Andrew Smock, and Caleb T Carr. Face [book] management: Self-presentation of political views on social media. *Communication Studies*, 66(5):549–568, 2015.

[37] Fritz Heider. Attitudes and cognitive organization. *The Journal of psychology*, 21(1):107–112, 1946.

[38] Fritz Heider. *The psychology of interpersonal relations*. Wiley, 1958.

[39] Sounman Hong and Sun Hyoung Kim. Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4):777–782, 2016.

[40] Daniel J Howard and Charles Gengler. Emotional contagion effects on product attitudes. *Journal of Consumer research*, 28(2):189–201, 2001.

[41] Norman P Hummon and Patrick Doreian. Some dynamics of social balance processes: bringing heider back into balance theory. *Social Networks*, 25(1):17–49, 2003.

[42] Herminia Ibarra. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative science quarterly*, pages 422–447, 1992.

[43] Myeongki Jeong, Hangjung Zo, Chul Ho Lee, and Yasin Ceran. Feeling displeasure from online social media postings: A study using cognitive dissonance theory. *Computers in Human Behavior*, 97:231–240, 2019.

[44] Wenjun Jiang and Jie Wu. Active opinion-formation in online social networks. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.

[45] Karl Kaltenthaler and William J Miller. The polarized american: Views on humanity and the sources of hyper-partisanship. *American Behavioral Scientist*, 56(12):1718–1734, 2012.

[46] Douglas T Kenrick, Norman P Li, and Jonathan Butner. Dynamical evolutionary psychology: Individual decision rules and emergent social norms. *Psychological review*, 110(1):3, 2003.

[47] Dennis Krebs. Empathy and altruism. *Journal of Personality and Social psychology*, 32(6):1134, 1975.

[48] Bibb Latané. The psychology of social impact. *American psychologist*, 36(4):343, 1981.

[49] Stéphane Laurens and Serge Moscovici. The confederate's and others' self-conversion: A neglected phenomenon. *The journal of social psychology*, 145(2):191–208, 2005.

[50] Jae Kook Lee, Jihyang Choi, Cheonsoo Kim, and Yonghwan Kim. Social media, network heterogeneity, and opinion polarization. *Journal of communication*, 64(4):702–722, 2014.

[51] Ke Li, Haiming Liang, Gang Kou, and Yucheng Dong. Opinion dynamics model based on the cognitive dissonance: An agent-based simulation. *Information Fusion*, 56:1–14, 2020.

[52] Lin Li, Anna Scaglione, Ananthram Swami, and Qing Zhao. Consensus, polarization and clustering of opinions in social networks. *IEEE Journal on Selected Areas in Communications*, 31(6):1072–1083, 2013.

[53] Bingjie Liu and S Shyam Sundar. Microworkers as research participants: Does underpaying turkers lead to cognitive dissonance? *Computers in Human Behavior*, 88:61–69, 2018.

[54] Yong Liu, Hongxiu Li, Xiaoyu Xu, Vassilis Kostakos, and Jukka Heikkilä. Modeling consumer switching behavior in social network games by exploring consumer cognitive dissonance and change experience. *Industrial Management & Data Systems*, 2016.

[55] Charles G Lord, Lee Ross, and Mark R Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098, 1979.

[56] Winter A Mason, Frederica R Conrey, and Eliot R Smith. Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. *Personality and social psychology review*, 11(3):279–300, 2007.

[57] David C Matz and Wendy Wood. Cognitive dissonance in groups: the consequences of disagreement. *Journal of personality and social psychology*, 88(1):22, 2005.

[58] Blake M McKimmie. Cognitive dissonance in groups. *Social and Personality Psychology Compass*, 9(4):202–212, 2015.

[59] Blake M McKimmie, Deborah J Terry, and Michael A Hogg. Dissonance reduction in the context of group membership: The role of metaconsistency. *Group Dynamics: Theory, Research, and Practice*, 13(2):103, 2009.

[60] Blake M McKimmie, Deborah J Terry, Michael A Hogg, Antony SR Manstead, Russell Spears, and Bertjan Doosje. I'm a hypocrite, but so is everyone else: Group support and the reduction of cognitive dissonance. *Group Dynamics: Theory, research, and practice*, 7(3):214, 2003.

[61] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[62] Daniel Mochon and Janet A Schwartz. Outrage drives facebook users to engage with ideology inconsistent political content. *ACR North American Advances*, 2019.

[63] Serge Moscovici et al. Toward a theory of conversion behavior. *Advances in experimental social psychology*, 13:209–239, 1980.

[64] Serge Ed Moscovici, Angelica Ed Mucchi-Faina, and Anne Ed Maass. *Minority influence.* Nelson-Hall Publishers, 1994.

[65] Daniel Nettle. Using social impact theory to simulate language change. *Lingua*, 108(2-3):95–117, 1999.

[66] Nyala Noë, Roger M Whitaker, Martin J Chorley, and Thomas V Pollet. Birds of a feather locate together? foursquare checkins and personality homophily. *Computers in Human Behavior*, 58:343–353, 2016.

[67] Samantha North, Lukasz Piwek, and Adam Joinson. Battle for britain: Analyzing events as drivers of political tribalism in twitter discussions of brexit. *Policy & Internet*, 2020.

[68] Michael I Norton, Benoit Monin, Joel Cooper, and Michael A Hogg. Vicarious dissonance: Attitude change from the inconsistency of others. *Journal of personality and social psychology*, 85(1):47, 2003.

[69] Andrzej Nowak, Jacek Szamrej, and Bibb Latané. From private attitude to public opinion: A dynamic theory of social impact. *Psychological review*, 97(3):362, 1990.

[70] John H Parmelee and Nataliya Roman. Insta-echoes: Selective exposure and selective avoidance on instagram. *Telematics and Informatics*, page 101432, 2020.

[71] Hafizh A Prasetya and Tsuyoshi Murata. A model of opinion and propagation structure polarization in social media. *Computational Social Networks*, 7(1):1–35, 2020.

[72] Fabio Sani. When subgroups secede: Extending and refining the social psychological model of schism in groups. *Personality and Social Psychology Bulletin*, 31(8):1074–1086, 2005.

[73] Fabio Sani, John Todman, and Judith Lunn. The fundamentality of group principles and perceived group entitativity. *Journal of Experimental Social Psychology*, 41(6):567–573, 2005.

[74] Joanne R Smith, Michael A Hogg, Robin Martin, and Deborah J Terry. Uncertainty and the influence of group norms in the attitude–behaviour relationship. *British Journal of Social Psychology*, 46(4):769–792, 2007.

[75] Jeff Stone and Joel Cooper. A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, 37(3):228–243, 2001.

[76] Laura M Strain. *Reducing Vicarious Dissonance: The Role of Group-Related Attributes and Ingroup Identification in Reduction Strategy Selection*. PhD thesis, Miami University, 2009.

[77] Sarah Tanford and Rhonda Montgomery. The effects of social influence and cognitive dissonance on travel purchase decisions. *Journal of Travel Research*, 54(5):596–610, 2015.

[78] Alexandru Topirceanu, Mihai Udrescu, Mircea Vladutiu, and Radu Marculescu. Tolerance-based interaction: A new model targeting opinion formation and diffusion in social networks. *PeerJ Computer Science*, 2:e42, 2016.

[79] Maykel Verkuyten and Kumar Yogeeswaran. The social psychology of intergroup toleration: A roadmap for theory and research. *Personality and Social Psychology Review*, 21(1):72–96, 2017.

[80] Sven Waldzus, Amelie Mummendey, Michael Wenzel, and Ulrike Weber. Towards tolerance: Representations of superordinate categories and perceived ingroup prototypicality. *Journal of Experimental Social Psychology*, 39(1):31–47, 2003.

[81] Clarisse Warren, Stephen Schneider, Kevin B Smith, and John R Hibbing. Motivated viewing: Selective exposure to political images when reasoning is not involved. *Personality and Individual Differences*, 155:109704, 2020.

[82] Gérard Weisbuch, Guillaume Deffuant, and Frédéric Amblard. Persuasion dynamics. *Physica A: Statistical Mechanics and its Applications*, 353:555–575, 2005.

[83] Roger M Whitaker, Gualtiero B Colombo, and David G Rand. Indirect reciprocity and the evolution of prejudicial groups. *Scientific reports*, 8(1):13247, 2018.

[84] Hywel TP Williams, James R McMurray, Tim Kurz, and F Hugo Lambert. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global environmental change*, 32:126–138, 2015.

[85] Matthew L Williams and Pete Burnap. Cyberhate on social media in the aftermath of woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, 56(2):211–238, 2016.

[86] Wendy Wood. Attitude change: Persuasion and social influence. *Annual review of psychology*, 51(1):539–570, 2000.

[87] Moran Yarchi, Christian Baden, and Neta Kligler-Vilenchik. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, pages 1–42, 2020.

[88] Sarita Yardi and Danah Boyd. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of science, technology & society*, 30(5):316–327, 2010.