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:


Publishers page: http://dx.doi.org/10.1109/TCSS.2021.3090833
<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\textsuperscript{1}, Gualtiero B. Colombo\textsuperscript{1}, Liam Turner\textsuperscript{1}, Yarrow Dunham\textsuperscript{2}, Darren K. Doyle\textsuperscript{3}, Eilish M. Roy\textsuperscript{3}, and Cheryl A. Giammanco \textsuperscript{4}

\textsuperscript{1}Crime and Security Research Institute, Cardiff University, Cardiff, UK
\textsuperscript{2}Department of Psychology, Yale University, New Haven, USA
\textsuperscript{3}Defence Science and Technology Laboratory, Human and Social Sciences Group, Porton Down, Salisbury, UK
\textsuperscript{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 individual’s 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 \( \text{sim}_{ij} \), is based on the proportion of fixed traits held in common, where:

\[
\text{sim}_{ij} = \frac{|\{k : f_k^i = f_k^j\}|}{n_f}
\]

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 \( t^*_ij = (1 - \text{sim}_{ij})t_i \). 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

\[
d^i = \sum_{j \in N_i} d_{ij} \cdot \text{sim}_{ij}
\]

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 it 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 counteract 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 that \( i \) chooses conviction level \( c_j \) from neighbour \( j \in N_i' \) is

\[
    d_{ij} + K \quad \sum_{j \in N_i'} (d_{ij} + K)
\]

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

\[
    \frac{\text{sim}_{ij} + K}{\sum_{k \in N_i} (\text{sim}_{ik} + K)}
\]

where \( K \) is a small constant. Upon selection, neighbour \( j \) is persuaded to adopt \( i \)'s conviction level \( c_i \) with probability \( \text{sim}_{ij} \). 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 \(-d_{ij} \geq -d_{ik}, \text{ for all } k \in N_i, k \neq i \). 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 rewrite 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 responses for agents 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 \( T_0, T_1, T_2 \). 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;
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 (\( \text{div}_{\text{avg}} \)), where an agent \( i \)'s neighbourhood diversity is defined as \( \text{div}_i = \sum_{j \in N_i} |c_i - c_j| \) and \( \text{div}_{\text{avg}} = \sum_{k \in V} \text{div}_k / |V| \);
- 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 (\( \text{sim}_{\text{avg}} \)) where neighbourhood similarity for agent \( i \) is defined as \( \text{sim}_{\text{avg}} = \sum_{j \in N_i} \text{sim}_{ij} / |N_i| \) and \( \text{sim}_{\text{avg}} = \sum_{k \in V} \text{sim}_{\text{avg}}^k / |V| \).

### 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 \( \text{mixed baseline} \)). 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 (\( \text{cd}^i \)) levels, taken from a sample of five runs for each starting network configuration.

#### IV.B. 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

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Main Response</th>
<th>( T_0 )</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>mixed baseline</td>
<td>( -3 )</td>
<td>( 0 )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>B</td>
<td>do nothing</td>
<td>( 3 )</td>
<td>( 3 )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>C</td>
<td>change self</td>
<td>( -3 )</td>
<td>( 3 )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>D</td>
<td>change other</td>
<td>( -3 )</td>
<td>( -3 )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>E</td>
<td>rewire</td>
<td>( -3 )</td>
<td>( -3 )</td>
<td>( -3 )</td>
</tr>
</tbody>
</table>
(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.

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
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,
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-conversion 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.

**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.
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’: 0.96%, ‘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, \( L_{100^4} \) indicates the 10,000-agent lattice and \( SF \) indicates the 10,000 node scale-free network starting configurations.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Starting Configuration</th>
<th>Scenario F</th>
<th>Scenario G</th>
<th>Scenario H</th>
<th>Scenario I</th>
<th>Scenario J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( L_{100} )</td>
<td>( L_{100^4} )</td>
<td>SF</td>
<td>( L_{100} )</td>
<td>( L_{100^4} )</td>
<td>SF</td>
</tr>
<tr>
<td>Average Cognitive Dissonance</td>
<td>1.926</td>
<td>1.914</td>
<td>1.891</td>
<td>0.551</td>
<td>0.593</td>
<td>0.537</td>
</tr>
<tr>
<td>Number of Alternative Conviction Levels</td>
<td>11.00</td>
<td>11.00</td>
<td>11.00</td>
<td>6.000</td>
<td>11.00</td>
<td>8.000</td>
</tr>
<tr>
<td>Average Conviction</td>
<td>5.250</td>
<td>5.250</td>
<td>5.250</td>
<td>5.854</td>
<td>5.842</td>
<td>5.433</td>
</tr>
<tr>
<td>Number of Alternative Tolerance Levels</td>
<td>6.000</td>
<td>6.000</td>
<td>6.000</td>
<td>1.000</td>
<td>5.000</td>
<td>5.000</td>
</tr>
<tr>
<td>Average Tolerance</td>
<td>3.190</td>
<td>3.190</td>
<td>3.190</td>
<td>0.000</td>
<td>0.078</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Clustering</td>
<td>0.000</td>
<td>0.000</td>
<td>0.988</td>
<td>0.074</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Assortativity by Conviction Levels</td>
<td>0.021</td>
<td>0.032</td>
<td>0.000</td>
<td>0.718</td>
<td>0.614</td>
<td>0.522</td>
</tr>
<tr>
<td>Average Local Diversity</td>
<td>2.446</td>
<td>2.403</td>
<td>2.303</td>
<td>0.554</td>
<td>0.572</td>
<td>0.485</td>
</tr>
<tr>
<td>Average Similarity</td>
<td>0.511</td>
<td>0.499</td>
<td>0.500</td>
<td>0.451</td>
<td>0.443</td>
<td>0.450</td>
</tr>
</tbody>
</table>

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

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

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 focused on

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.
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

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.
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, dominates 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=AGENT-BASED SIMULATION, IMPL=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=STUBBORNNESS (CANNOT BE PERSUADED), N/A=NOT APPLICABLE.

<table>
<thead>
<tr>
<th>Components</th>
<th>Primary Topic/Context</th>
<th>Cog. Diss.</th>
<th>Tolerance Polariation</th>
<th>Data Temporality</th>
<th>Key Contributions</th>
<th>Response to Dissonance</th>
<th>Relevant Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB OS SM AB</td>
<td>in-lab dissonance</td>
<td>exp. N/A N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B.C</td>
</tr>
<tr>
<td>[49]</td>
<td>in-lab dissonance</td>
<td>exp. N/A N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B.C</td>
</tr>
<tr>
<td>[63]</td>
<td>in-lab dissonance</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B.C</td>
</tr>
<tr>
<td>[57]</td>
<td>in-lab dissonance</td>
<td>exp. imp. imp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B.C,E,D,E</td>
</tr>
<tr>
<td>[59]</td>
<td>in-lab dissonance</td>
<td>imp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B.C,E</td>
</tr>
</tbody>
</table>

Clustering and more extreme levels of conviction, it is evident 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 converges 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
