Theoretical Framework and Model Settings

This project assumes that a small proportion of actors (10%) will adopt a set of corrections (a list of preferences) that is in sharp contrast to a list of older preferences. Instead of assuming that actors insist on and spread their preferences (one-time adoption), this project investigates how all actors who participate in the discussion will update their preferences in a manner that avoids cognitive dissonance under different network topologies.

The model assumes that people’s attitudes are subjects to the social influence from others (Becker, Porter and Centola 2019, DeMarzo, Vayanos and Zwiebel 2002) and to one’s own preferences and cognitive status (Page and Shapiro 2010, Taber and Lodge 2006). The social influence is modeled by one’s structural conditions, which assume people are subjected to social influences from local social interactions (local neighbors) and the society has a global structure of social interaction (network topologies). The cognitive status here is largely about motivated reasoning. In particular, the structural and cognitive conditions are as follows:

First, adaptations in Different structures can be model by testing attitude changes in different network topologies, such as

(1) scale-free networks (A type of networks with strong core-periphery structures. The cores not only occupy central positions in global topology but also have more local direct neighbors) (Barabási and Albert 1999);

(2) small-world networks (A type of network with weak core-periphery structures. Agents are embedded in different communities and these communities are randomly linked by some bridges. These bridges appear as cores because different communities rely on them to get connected) (Watts and Strogatz 1998);

(3) random regular networks (A type of network with no core-periphery structures. Agents not only occupy similar positions in global topology but also have the same number of direct neighbors) (Csardi and Nepusz 2006).

Second, the model captures motivated reasoning by assuming that agents update their preferences when the update does not increase their cognitive dissonant. Specifically, we consider motivated reasoning as two types of reasoning:

(1) weakly motivated reasoning, i.e., cognitive dissonance reduction. People are motivated to reduce dissonance. Dissonance is a psychological state of tension between any two cognitions, when considered by themselves, are obverse (Festinger 1957, Goldberg and Stein 2018, Shultz and Lepper 1996). Traditional social psychology research usually relies on theories of cold cognition to study dissonance reduction (Taber and Lodge 2006)

(2) strongly motivated reasoning, i.e., affective reasoning. People are still motivated to reduce dissonance, but the dissonance here is created by stimuli that would trigger strong affective responses (Bolsen, Druckman and Cook 2014, Taber and Lodge 2006). In this case, people feel strongly about issues and spend more time and cognitive resources to generate arguments. The attitudes are more likely to be polarized, strong, and sophisticated (Druckman et al. 2021, Taber and Lodge 2006).

I build on the associative diffusion model (AD) (Goldberg and Stein (2018)) to investigate that if a small proportion of actors at different structural positions adopt a correction, to what extent they would change the collective attitude towards the old misinformation. Although the AD model shares a similar cognitive assumption with this paper that actors adapt their preference depending on the level of cognitive consonance, it was originally developed to demonstrate endogenous cultural differentiation and does not pay close attention to preference change. Therefore, the original AD model does not precisely target the research purpose of examining collective preference change found in this paper, which departures from the AD model by making the five changes:

(1). For *correction of old misinformation*, while the AD model randomly assigns the value of all agents’ preferences from the same probability distribution (*V~U(-1,1)*), this paper assumes that agents’ preferences follow two different subjective probability distributions: a set of old, widespread preferences, *Vn*, and a set of new corrected preferences, *Vm*. Specifically, a large group of agents, *n*, form initial preferences, *Vn ~U(0,2*), and a small group of agents, *m,* adopt a set of corrections of preferences, *Vm~ U(-2,0*), which have the sharp contrast values with *Vn*.

(2) For *core-peripheral topologies*, this paper mainly discusses contagion networks with core-peripheral structures. In particular, this paper adopts the widely applied Barabási game to generate an original contagion scale-free network with strong core-peripheral structures (Barabási and Albert 1999). Agents observe the behaviors of their neighbors from the original contagion network and adapt their preferences. In particular, we further specify that the corrections may start from periphery agents or core agents, and compare these two different situations. For the purpose of comparison, we also simulate attitude changes in networks with weak core-peripheral structures and no core-peripheral structures.

(3) For updating preferences in *weak motivated reasoning*, the AD model has a latent assumption that agents display “centrism”—that the preference change, *∆v*, follows normal distribution *N(0,1)*, with the mean 0 and at the center of the original preferences, the uniform distribution *U(-1,1)*. This paper does not assume agents display centrism and consider them composed of both centrism and extremists in the case of weak motivated reasoning, and thus sets the *∆v* follows uniform distribution *U(-2,2)*.

(4) For updating preferences in *affective reasoning*, in contrast to the AD model’s latent assumption that agents display “centrism”, the affective reasoning argues that the preference change, *∆v*, is likely to become extreme (Druckman et al. 2021, Taber and Lodge 2006). Therefore, the preference change, *∆v*, is not likely to follow normal distribution *N(0,1)*. Instead, the *∆v* should reinforce the pre-adopted beliefs when exposed to opposing ideas (Bail et al. 2018). Therefore, when the update increase CS, then *∆v* follows uniform distribution *U(-2,2)*; when the update decrease CS for agent A, then *∆v* follows distribution *N(1,1)* ifvalue of *VA > 0,* and then *∆v* follows distribution *N(-1,1)* ifvalue of *VA < 0*

(5) Finally, the key output for our model is *the correction of collective judgment*, and we define “collective judgment” as the mean of all actors’ preferences. The extent of correction is the difference between initial collective judgment and the final connective judgment. The simulation ends when iterations reach given limits (1,000,000 times or higher) or agents’ preferences follow the same pattern, indicated by high group-level preference congruence. Goldberg and Stein (2018) measure the preference congruence between two agents A and B as “the absolute correlation between two agents’ preference vectors” (p. 912). To measure group-level preference congruence, they use the “mean absolute correlation between all pairs of agents’ preference vectors” (p. 912). The threshold for ending a game is group-level preference congruence > 0.99.