Group dynamics on multidimensional attitudes

Noah E. Friedkin*, Anton V. Proskurnikov†, Francesco Bullo‡

Abstract

The literature on attitudes toward objects includes seminal research indicating that individuals locate an object in a multidimensional appraisal space defined by the object's perceived degree of being good or bad, weak or strong, and passive or active. We advance this research in three ways. First, we generalize the information integration on an object with the inclusion of other individuals' displayed attitudes toward the object, and posit the existence of a dynamical system of information integration that generates a network of interpersonal influences on group members' object appraisals. Second, we show that this influence system entails a set of non-obvious and rarely violated constraints on individuals' settled appraisals. Third, with data collected in experiments on groups' appraisals of images of nine animals and two nations, Russia and North Korea, we report empirical findings that support the existence of this system and its predicted constraints on individuals' object appraisals.

Keywords:

Small groups; influence systems; multidimensional attitudes

Corresponding author:

Noah E. Friedkin, friedkin@soc.ucsb; PO Box 12936, Santa Barbara, CA 93107; 805-259-8389

*N.E. Friedkin is with the Department of Sociology and the Center for Control, Dynamical-Systems and Computation, at the University of California, Santa Barbara, USA e-mail: friedkin@soc.ucsb.edu

†A. V. Proskurnikov is with Politecnico di Torino, Turin, Italy and the Institute for Problems of Mechanical Engineering of the Russian Academy of Sciences, St. Petersburg, Russia e-mail: anton.p.1982@ieee.org

‡F. Bullo is with the Department of Mechanical Engineering and the Center for Control, Dynamical-Systems and Computation, at the University of California, Santa Barbara, USA e-mail: bullo@engineering.ucsb.edu

1 Introduction

Social psychology may be defined as the investigation of two dynamical systems and their interrelationship: (1) the individual as a dynamical system (iDS) that maps inputs onto an individual's attitude or behavior outputs,

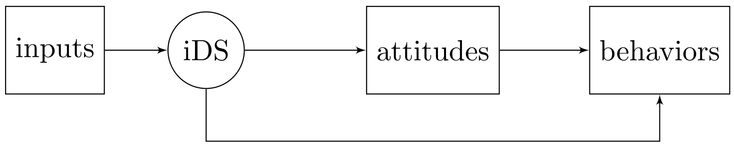


Figure 1: Individual Input-Output Function Machine

and (2) the group of individuals as a dynamical system (gDS) that maps inputs onto collective attitudes or behaviors

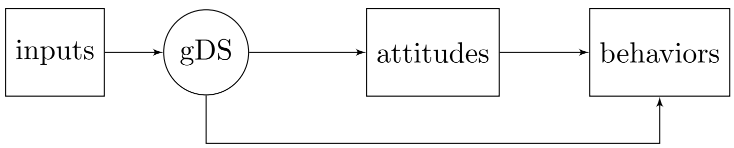


Figure 2: Group Input-Output Function Machine

Many of the investigative currents of social psychology are framed by these two dynamical systems. The iDS is the complex system of a human brain. The gDS is more open to alternative definitions; here we take it to be the group's set of attitudes, behaviors, and the social network structure and social processes that unfold on it. As such the gDS is also a complex system in which (i) the implications of network structure cannot be understood without precise definitions of the social processes that unfold on it, and (ii) the implications of a social process cannot be understood without precise definitions of the network structure of the group. Investigating the interrelationship of these two dynamical systems is a challenging endeavor. In sociology, it is associated with the qualitative premises of symbolic interaction theory (Blumer, 1969) that attend to the interactions among individuals in forming the meanings they ascribe to objects.

The paper is organized as follows. In Section 1.1, we review the evidence on attitude automaticity and, more generally, the evidence on rapid heuristic responses to objects. In Section 1.2, we review the evidence on the existence of an Euclidean appraisal space in which individuals locate their multidimensional attitudes toward objects. The above two lines of research are focused on independent individuals' iDS generated initial responses to objects. In Section 1.3, we attend to the gDS processing of individuals' initial attitudes. This is a natural generalization because, when individuals are nested in a communicating group, the information that is integrated on an object usually also includes the displayed appraisals of other individuals to the same object. We formalize this generalization with a novel application of the Friedkin-Johnson (Friedkin and Johnsen, 2011; Parsegov et al., 2017) information integration mechanism in which predictions of seven mechanistic constraints on individuals' attitude changes are derived. In Section 2, we describe the data collected from experiments on groups of human subjects with which we test these predictions, and in Section 3 we report the results of these tests. The novelty of these tests is that they pertain to the existence of a suite of implicit constraints on attitude changes that are implicated the standard global test of the association of observed and predicted settled attitudes. Thus, with these tests, we probe the validity of a deeper layer of non-obvious mechanistic constraints that are implicated in individuals' observed attitude changes than has heretofore been entertained.

1.1 Attitude automaticity

In this section, we address the evidence on attitude automaticity and, more generally, the evidence on rapid heuristic responses of the brain to objects. With advancements in cognitive science, it has become increasingly evident that the human brain automatically attends to and synthesizes available information, and that what constitutes "information" includes everything that is available to its sensory faculties and in memory. A manifestation of the evaluative activity of the mind is its automatic attitudinal responses to any perceived object, which are positive or negative cognitive orientations of particular intensity (Bargh and Ferguson, 2000; Bargh and Williams, 2006; Zajonc, 1980, 1998). Kahneman remarks that

> The evidence, both behavioral (Bargh, 1997; Zajonc, 1998) and neurophysiological (see, e.g., (LeDoux, 2000)), is consistent with the idea that the assessment of whether objects are good (and should be approached) or bad (and should be avoided) is carried out quickly and efficiently by specialized neural circuitry. Several authors have commented on the influence of this primordial evaluative system (here included in System 1) on the attitudes and preferences that people adopt consciously and deliberately (Epstein, 2003; Kahneman et al., 1999; Slovic et al., 2002; Wilson, 2000; Zajonc, 1998). (Kahneman, 2003, p. 701)

The available evidence also suggests that heuristically generated evaluative orientations are quickly translated by the mind into a variety of other forms of displayed attitudes and evaluations, e.g., subjective probabilities, rank orders, and monetary allocation preferences (Kahneman and Ritov, 1994; Leiserowitz, 2006; Slovic et al., 2002)

Intellective issues, which are issues with correct and incorrect positions, are typically dealt with by the same automatic mechanism. As the biases of individuals' responses to such issues have been elaborated, the assumption of ubiquitous rational calculation has become increasingly suspect. Kahneman's work on human judgment has accelerated this viewpoint (Kahneman et al., 1982; Kahneman, 2011). With the accumulating findings of the cognitive revolution in social psychology, indicating ubiquitous automatic mechanisms, it has become increasingly anomalous to assume logical reasoning and rational calculation. Although the slow hard work of logical reasoning and rational calculation may be done by some individuals, on some occasions in which such work is aroused, the quick work of an automatic mechanism appears to be a more generally accurate basis on which to construct models of the individual dynamical system (iDS) that generates appraisals of objects. Conscious mathematical reasoning falls into the class of effortful, slower, and deliberative analysis, and such reasoning may generate inescapable conclusions that override all individual differences among individuals who evaluate the mathematical reasoning. Apart from this special case, which requires expertise in mathematical analysis, deliberative analysis more typically takes the form of a less constrained debate of the merits of alternative courses of action. Simon's work (Simon, 1957) on bounded rationality presaged the cognitive revolution in its restriction of rational behavior to highly specialized circumstances in which there exist well-defined goals of action, limited optional courses of action, and known implications of choosing one course of action over another. Zajonc's work (Zajonc, 1980, 1998) also narrows the theoretical importance of conscious deliberation to an understanding of human cognitions and behaviors with the postulate that a substantial portion of such deliberation falls into the class of rationalizing the appraisals that the individual has already reached on the basis of automatic mechanisms.

The quick heuristic mechanisms of iDS information processing and integration are innate capabilities of a system that operates without a consciousness of the steps involved. A *fast conscious* process is an oxymoron in its implication of an awareness of the processing and integration of information that is generating an automatic appraisal of an object. Automatic appraisals of objects are based on the subconscious biologically-embedded heuristics of the iDS. The iDS generation of an initial attitudinal response to objects is subconscious, fast, and effortless. Were it not so, we would not have survived as a species.

1.2 Attitude structure universals in Euclidean semantic space

In this section, we address the evidence on the existence of a Euclidean appraisal space in which individuals' locate their multidimensional attitudes toward objects. The seminal evidence is presented by the psychologist C. E. Osgood and his collaborators in two books (Osgood et al., 1957, 1975). They describe the broad stroke grounding of their work as follows:

> Most social scientists would agree—that how a person behaves in a situation depends upon what that situation means or signifies to him. And most would also agree that one of the most important factors in social activity is meaning and change in meaning—whether it be termed "attitude," or "value," or something else again (Osgood et al., 1957, pp. 1). ... Of all the imps that inhabit the nervous system—that "little black box" in psychological theorizing—the one we call "meaning" is held by common consent to be the most elusive. Yet, again by common consent among social scientists, this variable is one of the most important determinants of human behavior. It therefore behooves us to try, at least, to find some kind of objective index. To measure anything that goes on within "the little black box" it is necessary to use some observable output from it as an index. ... We wish to find a kind of measurable activity or behavior of sign-using organisms which is maximally dependent upon and sensitive to meaningful states, and minimally dependent upon other variables. (Osgood et al., 1957, pp. 10-11)

For this index, they look to the particular words of a language that individuals assign to objects, and pursue a cross-cultural understanding these assignments. All languages contain a subset of *sense words* that directly refer to perceived dimensions of the physical properties of objects: e.g., vision-based words, touch-related words, olfactory-related words, taste-related words, auditory-related words. They investigate whether word assignments to objects define appraisal positions in a multidimensional Euclidean *semantic* space with a small number of culturally universal quantitative dimensions.

> Let us assume that there is some finite number of representational mediation reactions available to the organism and let us further assume that the number of these alternative reactions (excitatory or inhibitory) corresponds to the number of dimensions or factors in the semantic space. Direction of a point in the semantic space will then correspond to what reactions are elicited by the sign [object], and distance from the origin will correspond to the *intensity* of the reactions. (Osgood et al., 1957, p. 27)

Samples of individuals are presented with a sequence of objects and report their responses to each object on a set of semantic differential scales. A particular object is presented as an image or, more usually, a word (for example "tiger") that is a familiar object to all subjects. Each such scale is based on bipolar adjectives, for example, the [(bad) $-1 \ldots 0 \ldots 1$ (good)] scale. Tab. 1 is an example of an employed set of antonyms. For a given object, an individual recorded a position on each semantic differential. The language in which antonyms were presented varied depending on the native language of the individual. The first book (Osgood et al., 1957) is based on cross-cultural data collected on samples of individuals with three different native languages (English, Korean, or Japanese). The second book (Osgood et al., 1975) presents evidence from a massive cross-cultural undertaking on samples drawn from over 25 populations with different native languages.

The analysis of these data is concerned with the orthogonal factor structure of individuals' responses. The finding is that that individuals' object appraisals are reliably described as a location in a low dimensional 3D Euclidean space. They further find that the same three dimensions reliably arise and explain more of the variance of responses than all other detected dimensions. The most important of these dimensions is *evaluative*, the placement of an object on a cluster of related scales indicating the extent to which the object is good or bad. The second and third most important dimensions locate the object's *potency* (e.g., a cluster of related scales indicating the object's degree of strength or weakness) and *activity* (e.g., a cluster of related scales indicating the object's degree of activity or passivity). The evaluative factor explains double the variance of the potency and activity factors combined; and the latter two factors explain double the variance of all other factors combined. Thus, the meaning ascribed to a particular object for an individual is robustly an EPA position in a three-dimensional semantic space defined the dimensions of (E)valuation, (P)otency, and (A)ctivity. Remarkably, this three-dimensional space is cross-cultural and applies to a large variety of objects. It is, perhaps, not surprising that the EPA characterization of these three dimensions has a natural correspondence to appraisals of an object's potential threat.

1.3 Automatic social networks and influence systems

The above lines of research are focused on individuals' independent initial responses to objects. But when individuals are nested in a communicating group, the information that is integrated on an object usually includes the displayed attitudes of other individuals to the same object. Some individuals may see the object as malicious, active, and powerful while others may see it as malicious, active, and weak. The perceptions of other individuals' displayed attitudes and displayed changes

Table 1: Illustrations of employed semantic differentials

good-bad	strong-weak	passive-active
clean-dirty	big-little	noisy-quiet
nice-awful	powerful-powerless	courageous-timid
mild-harsh	hard-soft	intense-calm
beautiful-ugly	wild-tame	near-far
benign-hostile	sturdy-fragile	swift-slow

of attitudes are new stimuli that automatically trigger reactivations of individuals' information integration activity. Thus, we need to attend to the fact that individuals' displayed attitudes may influence the attitudes of other individuals, and that a social network of interpersonal influences is automatically constructed in which each individual's object appraisal may be influenced directly or indirectly by other individuals. A dynamical system of interpersonal influence is implicated that may reduce the group's heterogeneity of responses and generate a consensus appraisal of the object.

Research on social influence systems present many alternative specifications of mechanisms that alter individuals' attitudes. When tests of predictions have been conducted, the tests have generally focused on the association of individuals' predicted and observed settled attitudes, and more often than not find significant associations. *Thus, it is worthwhile to probe more deeply with tests of the predictions of the suite of mechanistic constraints on attitude changes that are intrinsic properties of a postulated mechanism.* Here we do just that on the application of the Friedkin-Johnsen model (Friedkin and Johnsen, 2011; Parsegov et al., 2017) to group dynamics on EPA multidimensional attitudes. We will assume that the group dynamics of object appraisals may be understood as epiphenomena of a shared iDS information integration mechanism that automatically incorporates the social information contained in other individuals' displayed object appraisals,

$$x_{id}(t+1) = a_{iid} \sum_{j=1}^{n} w_{ij} x_{jd}(t) + (1 - a_{iid}) x_{id}(0), \; t = 0, 1, 2, \ldots, \; d = 1, 2, 3, \tag{1}$$

for all i individuals in a group of n individuals. On each dimension d of the object, $x_{id}(0)$ is i's initial attitude, and $0 \leq a_{iid} \leq 1$ is i's level of openness to influence. The $0 \leq w_{ij} \leq 1$ is i's allocated relative weight to j's displayed attitudes at all times t, $\sum_{j=1}^{n} w_{ij} = 1 \; \forall i$. A rough analogue of this system is an iDS neural network architecture of information integration in which the neurons (nodes) flexibly adjust their action potentials. The group's $n \times n$ matrix of $a_{iid} w_{ij}$ weights generates an influence network with n nodes and a set of $i \xrightarrow{a_{iid} w_{ij} > 0} j$ arcs. In this network, each node's a_{iid} state regulates the $w_{ij} > 0$ influence of j's displayed appraisals of an object on i's appraisals. Thus, we define a network that may be adjusted by individuals' dimension specific levels of openness to interpersonal influence. If $a_{iid} = 0$, then i completely inhibits (discounts) its structural $w_{ij} > 0$ arcs and is stubbornly fixated on its dimension d initial appraisal. If $a_{iid} = 1 - w_{ii} = 1$, then i completely discounts its dimension d initial appraisal, and excites all of its $w_{ij} > 0$, $i \neq j$, arcs. An individual can be completely closed to influence on some appraisal dimensions and completely open to influence on other dimensions. Each object dimension is associated with a $n \times n$ matrix $\mathbf{A}_d\mathbf{W}$ in which $\mathbf{A}_d$ is a diagonal matrix with $0 \leq a_{iid} \leq 1$ for all i and $a_{iid} = 0$ for all $i \neq j$, and $\mathbf{W}$ is a row stochastic matrix of w_{ij} relative weights.

1.4 Predictions

From the Eq. 1 mechanism, we now derive a suite of predications on relationships of the iDS information integration mechanism with the gDS group dynamics system.

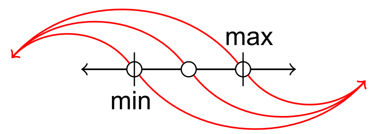
1.4.1 Prediction 1

If an initial consensus $\tilde{x}_d(0)$ exists on dimension d, $x_{1d}(0) = x_{2d}(0) = \ldots x_{nd}(0) = \tilde{x}_d(0)$, then the Eq. 1 mechanism predicts that it will be maintained: $\tilde{x}_{id}(t+1) = \tilde{x}_d(0)$ for all $t = 0, 1, \ldots$ is proved via induction on $t = 0, 1, \ldots$.

1.4.2 Prediction 2

All changes of appraisal are constrained to the min-max interval of the group's initial appraisals $x_{id}(0), i = 1, 2, \ldots n$, on each dimension d, that is, $[\min_i(x_{id}(0)), \max_i(x_{id}(0))]$. The initial displayed range of individuals' attitudes towards an object on a particular dimension of appraisal puts bounds on the possible emergent attitudes, and the group's min-max initial appraisals on each of the dimensions create a cognitive box that is a constraining group-specific appraisal subspace in which all final appraisals are predicted to reside. This prediction is also proved via induction on $t = 0, 1, \ldots$ and follows from the convex combination property of Eq. 1, that is, $a_{iid} \sum_{j=1}^{n} w_{ij} + (1 - a_{iid}) = 1$. Fig. 3 illustrates possible violations of this predicted constraint.

Figure 3: Illustrations of violations of the prediction that all changes of appraisal are constrained to the min-max interval of the group's initial appraisals on each dimension d.



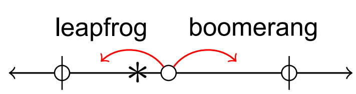
1.4.3 Prediction 3

The Eq. 1 mechanism predicts that any change of an individual's appraisal is a movement toward the individual's weighted average, $\sum_{j=1}^{n} w_{ij}x_{jd}(t)$, of the group's attitudes on each dimension d. The equilibrium equation of each individual's settled multidimensional attitude changes is given by

$$x_{id}(\infty) - x_{id}(0) = a_{iid}\left(\sum\nolimits_{j=1}^{n} w_{ij}x_{jd}\left(\infty\right) - x_{id}\left(0\right)\right). \quad (2)$$

The individual's change $x_{id}(\infty) - x_{id}(0)$ is proportional to the change $\sum_{j=1}^{n} w_{ij}x_{jd}(\infty) - x_{id}(0)$, and the individual's $0 \leq a_{iid} \leq 1$ level of openness to interpersonal influence on the appraisal of dimension d is the proportionality factor. Eq. 2 predicts (i) that $|x_{id}(\infty) - x_{id}(0)| \leq |\sum_{j=1}^{n} w_{ij}x_{jd}(\infty) - x_{id}(0)|$ and (ii) that the sign $(-, 0, +)$ of the observed appraisal change $x_{id}(\infty) - x_{id}(0)$ is identical to the sign of $\sum_{j=1}^{n} w_{ij}x_{jd}(\infty) - x_{id}(0)$. (iii) Thus, there should be no instances of out-of-bounds $a_{iid} < 0$ boomerang movements (away from the weighted average attractor $\sum_{j=1}^{n} w_{ij}x_{jd}(\infty)$), or $a_{iid} > 1$ leapfrog movements (over the weighted average attractor $\sum_{j=1}^{n} w_{ij}x_{jd}(\infty)$). Fig. 4 illustrates possible violations of this predicted constraint.

Figure 4: Illustrations of violations of the prediction that all changes of appraisal are movements toward the individual's weighted average $(*)$ of the group's attitudes on each dimension d.



1.4.4 Prediction 4

The Eq. 1 mechanism predicts that an appraisal change event will occur on dimension d if $0 < a_{iid} \leq 1$ and $0 \leq w_{ii} < 1$. In general, the magnitude of an appraisal change depends on the self-absorption factor $0 \leq a_{iid}(1 - w_{ii}) \leq 1$ that is implicated in the mechanism. Maximal movement toward the weighted average attractor is predicted if i's self-absorption level is minimal $a_{iid}(1 - w_{ii}) = 1$, that is, if $a_{iid} = 1 \wedge w_{ii} = 0$. No movement toward the weighted average attractor is predicted if i's self-absorption level is maximal $a_{iid}(1 - w_{ii}) = 0$, that is, if $a_{iid} = 0 \vee w_{ii} = 1$. Thus,

$$|x_{id}(\infty) - x_{id}(0)| = \begin{cases} 0, & \text{if } a_{iid}(1 - w_{ii}) = 0 \\ > 0, & \text{otherwise} \end{cases} \tag{3}$$

1.4.5 Prediction 5

The Eq. 1 mechanism predicts which of two types of attitude changes may occur: a change that preserves the sign of the attitude (an increase or decrease in the magnitude of the attitude with no change in sign), and a change that flips the sign of the attitude. The mechanism predicts the conditions under which an initial positive $x_{id}(0) > 0$ attitude changes to a negative $x_{id}(\infty) < 0$ attitude or vice versa. A change of the sign on dimension d is predicted if and only if $x_{id}(\infty)/x_{id}(0) < 0$, or, equivalently, if the inequality holds as follows

$$\frac{x_{id}(\infty)}{x_{id}(0)} = a_{iid} \frac{\sum_{j=1}^{n} w_{ij} x_{jd}(\infty)}{x_{id}(0)} + (1 - a_{iid}) < 0. \tag{4}$$

1.4.6 Prediction 6

The Eq. 1 mechanism predicts conditions under which a group consensus will generated. A consensus may be reached that is one of three types: it may be a compromise appraisal in the range of the initial appraisals that is not any of the group's initial appraisals, or it may be a settlement on one of the group's initial appraisals (minimum, maximum, or other initial appraisal), or an anomalous breaching consensus that is more extreme than the group's min-max initial appraisals on the dimension. Given two or more individuals in a group with disagreeing initial appraisals on object appraisal dimension d, one of two mutually exclusive network topological conditions (C_1, C_2) must be satisfied to reach an appraisal consensus in the min-max interval of its initial appraisals *for all possible arrays of heterogeneous initial appraisals.* These (C_1, C_2) conditions are based on structural features of the influence network $\mathcal{G}_d$ of the group on dimension d that is defined by the set of influence arcs $i \xrightarrow{a_{iid} w_{ij} > 0} j$. If C_1 is satisfied, then the prediction is that the group will reach a compromise consensus that is unlikely to be any of the group's initial appraisals. This condition is satisfied if $\mathcal{G}(d)$ is a strongly connected aperiodic network in which all individuals are maximally open to influence on dimension d, $a_{iid} = 1 \; \forall i$. It is strongly connected if every individual directly or indirectly influences all other individuals, that is, there exists at least one path of $u \xrightarrow{a_{iid}w_{ij}>0} v$ arcs from every i to every j in the network. Given a strongly connected network, the existence of at least one $0 < w_{ii} < 1$ suffices to secure aperiodicity. If C_2 is satisfied, then a consensus on one of the initial appraisals will be reached. This condition is satisfied if (i) the group has just one individual with $a_{iid} < 1$, and all remaining individuals are influenced by him/her (that is, the corresponding node in the graph is globally reachable), or (ii) the group has $k > 1$ individuals with $a_{iid} < 1$, their initial appraisals are identical, and each of the remaining $n - k$ individuals is influenced by at least one of them.

1.4.7 Prediction 7

The Eq. 1 mechanism predicts a significant *linear* correspondence of expected final appraisals $\hat{x}_{id}(\infty)$ with observed final appraisals $x_{id}(\infty)$ under conditions of an influence network $\mathcal{G}_d$ topology that are consistent with the convergence of predicted appraisals to a steady state (see (Friedkin and Johnsen, 2011; Parsegov et al., 2017) on these conditions). The mechanism's conditions of convergence to a steady state of predicted appraisals are quite broad: for example, it may fail to converge if the assumption of $0 \leq a_{iid} \leq 1$ is violated with $a_{iid} > 1$ or $a_{iid} < -1$ for some i, but it may converge if $| a_{iid} |< 1$ for all i.

2 Data and Methods

With data collected from experiments on groups of human subjects in the U.S., we test the 1-7 predictions of the mechanism. These data include human subjects' appraisals of two nations (Russia and North Korea) and the Fig. 1 images of nine animals. The units of analysis are $3,882 = 30 \times 3 \times 12$ (i, d) appraisal occasions in which individual i locates the object on attitude scales corresponding to the object's perceived degree of friendliness or hostility ($d = 1$), strength or weakness ($d = 2$), and passivity or activity ($d = 3$). In each experiment 1-2 (Russia, North Korea), there are $107 \times 3 = 321$ appraisal occasions. In experiment 3 we repeated an experiment on North Korea given its dynamic status in the news on its development of nuclear weapons during the 2018-2019 period of our data collection. There are $108 \times 3 = 324$ appraisal occasions. In the 9 pooled animal image experiments 4.1-4.9, there are $108 \times 3 \times 9 = 2,916$ appraisal occasions. On each occasion, the collected data are an individual's report of initial independent appraisals $x_{id}(0)$, post-discussion final appraisals $x_{id}(\infty)$, and the relative subjective influence weights $0 \leq w_{ij} \leq 1$,

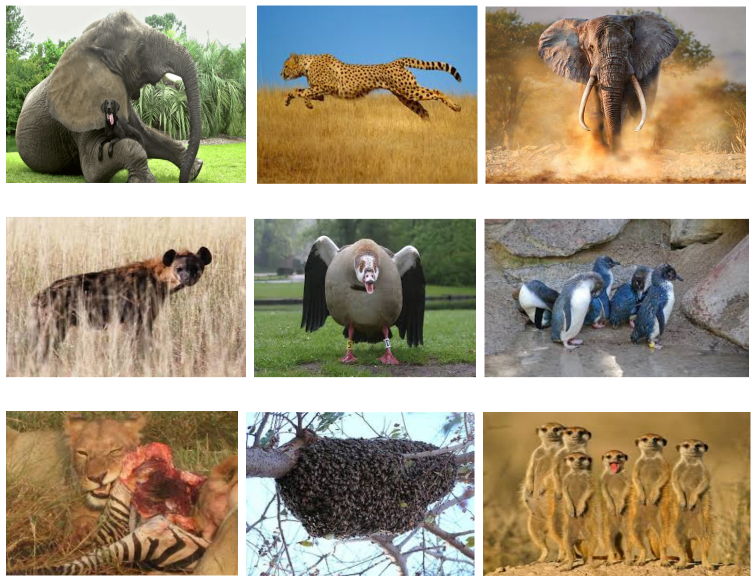
$\sum_{j=1}^{n} w_{ij} = 1$, allocated by i to each group member in determining i's appraisals. Subjects were instructed that reaching consensus is desirable but not required.

Experiments 1-2 collected data on 107 subjects nested in 30 groups with 3-4 members. In Experiment 1 on Russia, the following questions were posed. What are your appraisals of the current Russian government's posture toward the United States? Subjects' appraisals were expressed as a number between -100 and 100. How good or bad are its intentions and goals towards the United States: How active or inactive is it in pursuing its intentions and goals towards the United States? What is its level of capability of presenting a clear and present danger to the United States? In Experiment 2 on North Korea, the following questions were posed. How certain are you that each of the following 3 statements are true? Subjects' appraisals were expressed as a number between 0 and 100%. North Korea is developing intercontinental nuclear ballistic missiles. North Korea's weapons of mass destruction are a clear and present danger to the region and to the United States. A preemptive military action, which demonstrates U.S. capacity to intercept North Korean ballistic missiles in flight, or which incapacitates North Korean missiles launches, is justified.

Experiments 3 and 4.1-4.9 collected data on a different sample of subjects: 108 subjects nested in 30 groups with 3-4 members. In Experiment 3 on North Korea, the following questions were posed. Subjects' appraisals were expressed as a number between 0 and 100%. How certain are you that each of the following 3 statements are true? Do you believe that North Korea has the capacity (the ability or power) to harm the U.S., or do you believe that North Korea does not have such a capacity? Do you believe that North Korea's attention or activity is currently focused on the U.S., or it is oriented to elsewhere? Is North Korea currently indicating an aggressive attitude toward the U.S., or is it not currently indicating such an attitude? In Experiments 4.1-4.9, we presented subjects with images of the nine animals shown in Fig. 2. The scenarios associated with these images involved group encounters with these animals. Subjects' appraisals were expressed as a number between 0 and 100%. On each image, we posed the questions: Does this animal (or do these animals) have the capacity (the ability or power) to harm you, or does it (or do they) not have such a capacity? Is this animal's (or are these animal's) attention or activity currently focused on you, or is it (or are they) oriented elsewhere? Is this animal (or are these animals) currently indicating an aggressive attitude toward you (readiness to harm you), or is it (or are they) not currently indicating such an attitude toward you?

Note that we varied the measurement scales of the attitudes. All the measurement scales are standardized to $[-1, +1]$ interval scales in which positive attitudes are associated with high threat levels (bad, strong, active) and negative attitudes are associated with low threat levels (good, weak, inactive). The $[-100, +100]$ scales were transformed to $[-1, +1]$ with $\frac{1}{100}[-100, +100]$, and the $[0, 100]$ scales were transformed to $[-1, +1]$ with $-1 + \frac{2}{100}[0, 100]$.

Figure 5: The animal images evaluated by subjects.



2.1 Descriptive features of the data

In Fig. 6(A), we find that most individuals have a unique initial object appraisal position, $\mathbf{x}_i(0) = [x_{i1}(0), x_{i2}(0), x_{i3}(0)]$, that is not shared by any other individual in each of the 12 experiments. The usual effect of an interpersonal influence system is the elimination or reduction of displayed initial disagreements among individuals. In Fig. 6(B), we find that an exact consensus on all three dimensions occurred in the majority of the 30 groups in each experiment. In the aggregate, a group consensus on all three dimensions of appraisal occurred in 72.5% of the $360 = 30 \times 12$ possible occasions for a group consensus. Fig. 6(C) gives the frequency histograms of the observed appraisal changes. The mode of each distribution are individuals who did not alter their appraisals on a particular dimension of the object. The high relative frequency of individuals with no change of appraisal does not necessarily imply a high rate of failures to reach consensus. In Tab. 2 we find that within-group appraisal variability declined, and the proportion of total between-group appraisal variance increased on each dimension. Fig. 7 shows individuals' initial appraisal positions and their groups' mean final appraisal positions. The influence systems of the groups operated to differentiate groups' appraisals of the same object. In Tab. 3, we find that individuals often alter their appraisal on at least one dimension, and do not alter their appraisals on at least one other dimension.

Figure 6: (A) Most individuals' have a unique initial object appraisal position $\mathbf{x}_i(0) = [x_{i1}(0), x_{i2}(0), x_{i3}(0)]$, that is not shared by any other individual in each of the 12 experiments. (B) An exact consensus on all three dimensions occurred in the majority of the 30 groups in each experiment. (C) The mode of the frequency distribution of appraisal changes are individuals who did not alter their appraisals on a particular evaluative dimension of the object.

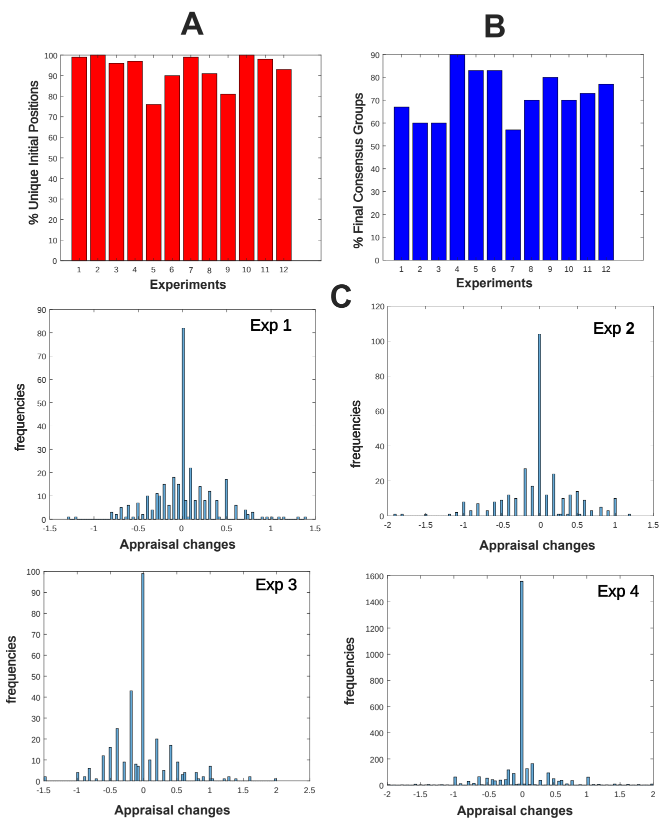


Table 2: Within-group appraisal variability declined, and the proportion of total between-group appraisal variance increased on each dimension. BGd are the percentages of the initial or final appraisal total variance that is between-groups on a particular appraisal dimension.

Experiments	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Initial Btw-Grp												
BGd-1	45%	42%	42%	57%	39%	42%	26%	40%	27%	37%	27%	30%
BGd-2	31%	46%	42%	33%	41%	32%	34%	26%	36%	22%	32%	31%
BGd-3	22%	35%	35%	29%	24%	42%	46%	51%	43%	28%	39%	56%
Final Btw-Grp												
BGd-1	86%	99%	91%	94%	92%	92%	88%	93%	24%	94%	92%	94%
BGd-2	94%	86%	92%	97%	85%	72%	58%	95%	91%	92%	97%	70%
BGd-3	87%	78%	93%	98%	69%	79%	95%	96%	97%	32%	86%	97%

Figure 7: Individuals' initial $(x_{i1}(0), x_{i2}(0), x_{i3}(0))$ appraisal positions ($*$) and their groups' mean final appraisals ($\cdot$) in each of the experiments.

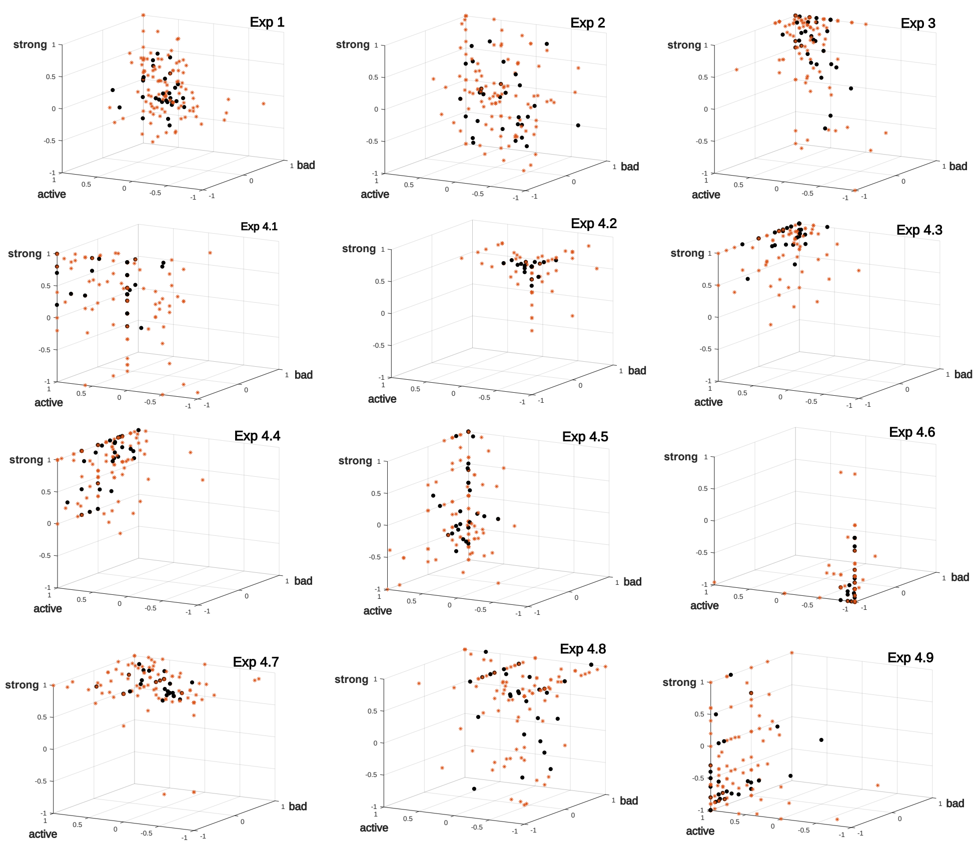


Table 3: Frequency counts of individuals' changes of attitude on each dimension: d-1 = d-2 = d-3 = 1 indicates individuals who altered their attitudes on all three object dimensions, and d-1 = d-2 = d-3 = 0 indicates individuals who did not alter their appraisals on any of the three object dimensions.

d-1	d-2	d-3	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
1	1	1	50	39	36	12	8	13	13	19	2	8	18	15
0	1	1	13	11	6	31	5	1	2	5	15	0	7	7
1	0	1	10	15	14	4	8	19	29	21	1	1	10	31
1	1	0	14	12	10	10	14	23	11	4	1	62	24	2
0	0	1	7	10	9	25	14	4	9	26	43	3	5	18
0	1	0	10	10	16	16	5	3	4	2	12	10	15	3
1	0	0	3	4	7	1	15	20	29	14	0	14	17	12
0	0	0	5	6	10	9	39	25	11	17	34	10	12	20

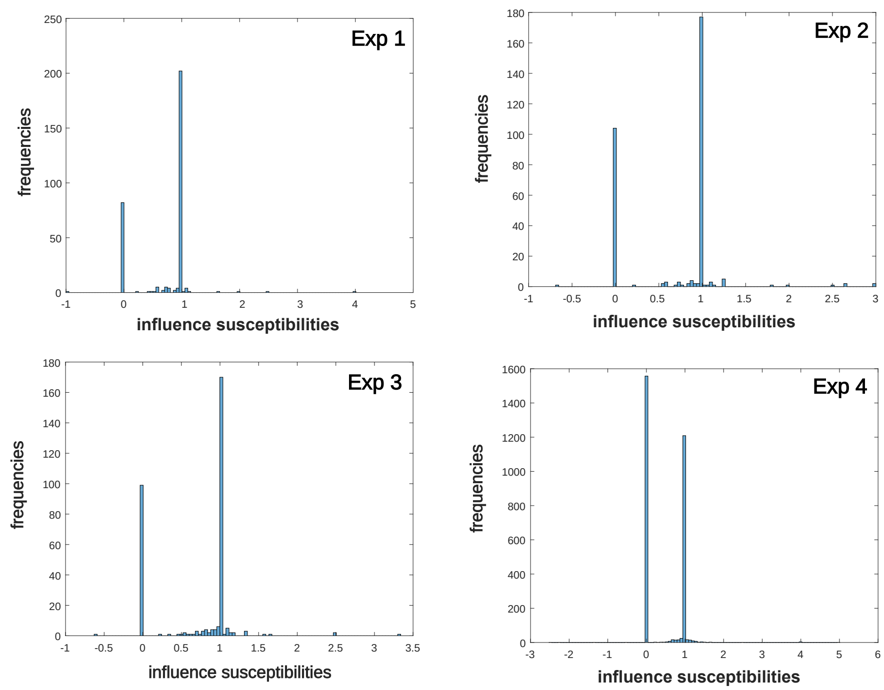
2.2 Derivation of a_{iid} susceptibilities to influence

To derive individuals' susceptibilities to influence, we distinguish two cases. In the first case, $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) = x_{id}(0)$. This may happen, in particular, if individual i is closed to social influence $w_{ii} = 1$ and $w_{ij} = 0$ for all $j \neq i$, or if there is an initial group consensus, or if a consensus has been reached on i's initial $x_{id}(0)$ appraisal. In this situation, the mechanism predicts that the appraisal of individual i remains unchanged $x_{id}(\infty) - x_{id}(0) = 0$. The coefficient a_{iid} in such a situation is not unique, and we formally define it as a_{iid}=0. We find that occasions of $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) - x_{id}(0) = 0$ exist in all 12 experiments 1-3 and 4.1-4.9 (48, 71, 92, 143, 193, 156, 127, 137, 210, 124, 121, 136, respectively), and confirm that $x_{id}(\infty) - x_{id}(0) = 0$ without exception. In the second case, $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) \neq x_{id}(0)$ and an individual's level of susceptibility is uniquely determined from

$$a_{iid} = \frac{x_{id}(\infty) - x_{id}(0)}{\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) - x_{id}(0)}. \tag{5}$$

Figure 8 gives the distribution of the derived a_{iid} susceptibility values. Individuals' susceptibilities to influence are predominately either maximal $a_{iid} = 1$ (complete openness to influence) or minimal $a_{iid} = 0$ (complete closure to influence).

Figure 8: Derived susceptibilities to influence. The basis of the percentages are the $107 \times 3 = 321$ occasions in experiments 1-2, and the $108 \times 3 = 324$ occasions in experiments 3 and 4.1-4.9. The susceptibilities in the animal image experiments 4.1-4.9 have been pooled.



3 Results on the Tests of Predictions

Our first prediction is that if an initial consensus exists on a particular object appraisal dimension, then it will be maintained. In Tab. 3, we find that this prediction is confirmed in 98.6% (all but 3) of the 216 observed occasions of an initial consensus.

Table 4: Testing the prediction that an initial consensus will be maintained.

	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Test Occasions	1	4	10	15	33	24	19	16	41	18	14	21
Confirmations	100%	75%	90%	100%	100%	100%	100%	100%	100%	100%	100%	95%

Our second prediction is that all changes of appraisal will be constrained to the min-max interval of the group's initial attitudes on each dimension. In Tab. 5, we find that this prediction is confirmed in 98.1% of the 3,882 occasions in which this constraint might be violated. The maintenance of an initial consensus is also a corollary of this constraint because, in this case, the group's min-max initial attitudes are identical, and any change of attitude is a violation of the group min-max initial attitude constraint.

Table 5: Testing the prediction that final appraisals are constrained to the min-max range of the group's initial attitudes on each dimension.

	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Test Occasions	321	321	324	324	324	324	324	324	324	324	324	324
Confirmations	99%	95%	96%	98%	99%	99%	99%	100%	100%	99%	98%	99%

Our third prediction is on the existence of weighted-average attractor constraints ($\mathbf{Cx}, \mathbf{Cy}, \mathbf{Cz}$) on attitude changes. The **Cx** constraint is that $|x_{id}(\infty) - x_{id}(0)| \leq |\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) - x_{id}(0)|$. The **Cy** constraint is that the sign $(-, 0, +)$ of the observed appraisal change $x_{id}(\infty) - x_{id}(0)$ is identical to the sign of $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty) - x_{id}(0)$. Thus, the **Cz** constraint is that there should be no instances of out-of-bounds $a_{iid} < 0$ boomerang movements in which i moves away from the weighted-average attractor $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty)$, or $a_{iid} > 1$ leapfrog movements in which i jumps over the weighted-average attractor $\sum_{j=1}^{n} w_{ij} x_{jd}(\infty)$. In Tab. 6, the finding is that, in the aggregate of the 3,882 response occasions, 86% satisfy all three predictions. Note that the violations of the $0 \leq a_{iid} \leq 1$ constraint are concentrated on instances of $a_{iid} > 1$, which indicates that most of these violations involve leapfrog movements toward greater perceived threat.

Table 6: Testing the predictions of weighted-average attractor constraints ($\mathbf{Cx}, \mathbf{Cy}, \mathbf{Cz}$) on attitude changes.

Cx	**Cy**	**Cz**	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
1	1	1	265	263	254	299	293	295	269	275	298	283	278	273
0	1	1	11	6	15	4	10	8	10	6	6	7	7	15
1	0	1	34	33	32	13	19	16	32	28	19	23	30	30
0	1	0	10	18	22	8	2	5	12	14	1	11	8	5
1	0	0	1	1	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
$a_{iid} > 1$			10	18	22	8	2	5	12	14	1	11	8	5
$a_{iid} < 0$			1	1	1	0	0	0	1	1	0	0	1	1

Our forth prediction is that an attitude change event will occur if and only if $0 < a_{iid} \leq 1$ and $0 \leq w_{ii} < 1$. In Tab. 7, we find that this condition is confirmed in 90% of the 2,015 appraisal change occasions in which this constraint might be violated.

Table 7: Testing the predictions of self-absorption constraints on appraisal changes.

	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Test Occasions	239	217	200	168	112	152	165	159	95	177	173	158
Confirmations	94%	88%	86%	95%	94%	93%	90%	87%	80%	88%	94%	92%

Our fifth prediction specifies the conditions under which an initial positive $x_{id}(0) > 0$ appraisal (some threat exists) changes to a negative $x_{id}(\infty) < 0$ appraisal or vice versa. Two types of changes of appraisals may occur: a change that preserves the sign of the appraisal (an increase or decrease in the magnitude of the appraisal with no change in sign), and a change that flips the sign of the appraisal. We find low frequencies of sign changes. In the aggregate of the 3,882 appraisal occasions, 6.5% involved changes of positive $x_{id}(0) > 0$ to negative $x_{id}(\infty) < 0$ or changes of negative $x_{id}(0) < 0$ to positive $x_{id}(\infty) > 0$. In Tab. 8, we find that the predictions of sign changes are always confirmed.

Table 8: Testing the predictions of appraisal sign changes.

	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Test Occasions												
$P \to N$	20	22	4	2	2	2	10	6	4	30	23	13
$N \to P$	15	12	7	14	0	12	10	15	0	8	16	3
Confirmations	100%	100%	100%	100%	100%	100%	100%	100%	100	100%	100%	100%

Our sixth prediction specifies the conditions of consensus formation. Given some initial appraisal heterogeneity, our findings in Tab. 9 are (i) in the aggregate of the 205 test occasions that satisfy the C_1 condition, 93% reached a compromise consensus in the min-max interval of initial appraisals that is not any of the initial appraisals, (ii) in the aggregate of the 466 test occasions that satisfy the C_2 condition, 94% reached a consensus on an initial appraisal, and (iii) in the aggregate of the 186 test occasions that satisfy neither the C_1 or C_2 conditions, 100% failed to reach consensus in the min-max interval of initial appraisals.

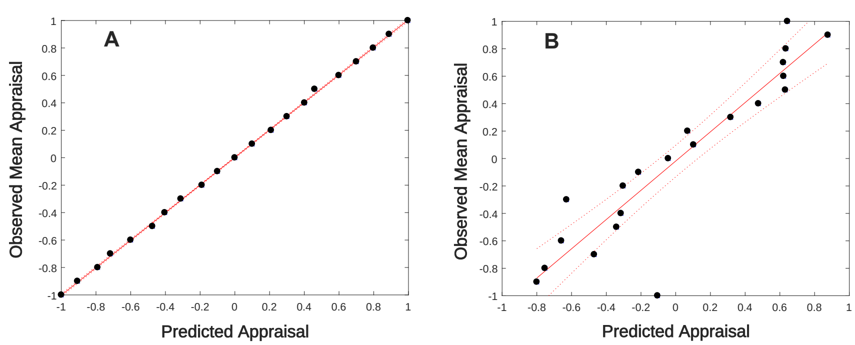
Table 9: Testing the predictions of consensus formation.

	1	2	3	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Compromise cns												
C_1 Test Occasions	32	21	17	17	13	19	14	17	4	19	14	18
Confirmations	94%	81%	88%	82%	92%	100%	100%	94%	100%	100%	93%	100%
Cns on an initial												
C_2 Test Occasions	30	44	35	51	36	32	37	38	30	31	48	34
Confirmations	100%	100%	100%	97%	100%	100%	97%	93%	100%	96%	100%	89%
No Cns												
$\sim (C_1 \vee C_2)$ Test Occasions	22	21	26	5	9	11	17	20	9	17	16	13
Confirmations	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Finally, we test the mechanism's prediction of final appraisals $x_{id}(\infty)$ that have a significant *linear* correspondence with the observed final appraisals under the condition of $0 \leq a_{iid} \leq 1$ for all i. The following important approximation generally applies to obtain the mechanism's predicted appraisals $\hat{\mathbf{x}} = (\mathbf{I} - \alpha \mathbf{A}\mathbf{W})^{-1}(\mathbf{I} - \alpha \mathbf{A})\mathbf{x}(0)$ for $\alpha \to 1$ for each dimension-specific $\mathbf{A}$ and $\mathbf{x}(0)$.

Among the 3,882 total occasions, 97% have derived $0 \leq a_{ii} \leq 1$. We find a marked difference in the exactitude of predictions depending on whether i is in a group with not all $a_{iid} = 1$, which comprise 78% of the occasions, or in a group with all $a_{iid} = 1$, which comprise 22% of the occasions. In the 2,759 occasions of i in a group with not all $a_{iid} = 1$, the linear regressions $R^2 = 0.995$, and near exact predictions occur in 97% of the occasions. Fig. 9(A) evaluates the correspondence of 21 bins of predicted appraisal values and the means of the observed values that are associated with each of them. In contrast, we find that in the 784 occasions of i in a group with all $a_{iid} = 1$, that is a group where all individuals are completely open to influence, the linear regressions $R^2 = 0.504$. The predictions are noisy with substantial variation in the distribution of observed values for each predicted value. Fig. 9(B) evaluates the correspondence of 21 bins of predicted appraisal values and the means of the observed values that are associated with each of them. The estimated means of the regression fall on a line passing through an intercept $\beta_0 = -0.0185$ (p = 0.741) that is not significantly different from 0 and a slope $\beta_1 = 1.066$ (p = 4.4597e-09) that includes 1 in its 95% CI. Thus, the expected observed values track along the line that is predicted by mechanism. The difference between Figs. 9 (A) and (B) may be understood as follows. When i is in a group with not all $a_{iid} = 1$ and condition C_2 is satisfied, then the prediction is insensitive to the network's w_{ij} values, that is, only the C_2 topology is determinative. When i is in a group with all $a_{iid} = 1$, then predicted appraisals are sensitive to any measurement errors on the network's w_{ij} values.

Figure 9: Evaluation of the correspondence of observed and mechanism predicted settled appraisals. (A) The correspondence is nearly deterministic for individuals who are nested in a group with not all $a_{id} = 1$. (B) The correspondence is noisy with expected observed values that track along the line that is predicted by mechanism for individuals who are nested in a group with all $a_{id} = 1$.



4 Discussion

In this article we have investigated the group dynamics that may alter individuals' initial appraisals of encountered objects. The hypothesis investigated is that the group dynamics of object appraisals may be understood as epiphenomena of a shared information integration mechanism that automatically incorporates the social information contained in other individuals' displayed object appraisals. An aspect of this hypothesis is that such a mechanism generates network structures of interpersonal influences on the basis of which both direct and indirect influences on appraisals occur. Thus, the implications of the information integration activity of individuals depends on structural features of the social influence networks that the mechanism created, and the implications of structural features of the networks depend on the information integration mechanism that created the networks. The contributions of this investigation are twofold. Its substantive contribution is the advancement of the EPA paradigm on individuals' multidimensional object appraisals. This paradigm has focused on individuals' independent appraisals. We generalize the paradigm to allow an influence system in which other individuals' displayed appraisals may alter individuals' appraisals. This is a natural generalization in which the information integration mechanism that is processing sensory inputs on an object's features and behaviors include sensory inputs of other individuals' displayed appraisals. Its theoretical contribution is the demonstration that a postulated model of an information integration mechanism sets up an influence system that implicitly involves a suite of testable predictions with a deeper bearing on the validity of a mechanism than is afforded by a test of the association of individuals' observed settled appraisals and predicted settled appraisals. For example, although the Fig. 9 plots of the linear association are an important feature of evaluating the validity of the postulated mechanism, the other presented tests have disgorged additional foundations of mechanistic validity that point to the existence of constraints on appraisal changes. Our findings on the tests of these predictions suggest that the group dynamics of object threat appraisals are subject to a set of general nonobvious constraints.

Social networks abound, their $i \to j$ arcs may be defined in various ways, and various processes may enfold on them. Rather than starting with a given social network, we start with a mechanism of individual information integration that automatically generates a social network when individuals are nested in a communicating group. This network is assembled by each individual's bundle of $i \xrightarrow{a_{iid}w_{ij}} j$ arcs of i's allocation of weights to themselves and others. The collection of these bundles create a social network of direct interpersonal influences. We do not investigate the suite of variables that may affect individuals' a_{iid} levels of openness-closure to interpersonal influence or their w_{ij} allocations of relative influence. The theoretical focus is on the implications of the mechanism for given set of measures of the mechanism's constructs. It may be that the structure of the influence network is also constrained by fundamental rules, and we have not investigated the existence of such constraints. Our approach and findings suggest the existence of an interesting nexus of neuroscience and social network science. In particular, we point to the possible linkage of our findings and conclusions with recent work on the social brain hypothesis (Dunbar, 2002, 2007; Dunbar and Shultz, 2007; Dunbar, 1998, 2009; Falk and Bassett, 2017; Sallet et al., 2011; Kanai et al., 2011). This hypothesis, proposed by British anthropologist Robin Dunbar, broadly deals with idea that the evolution of human intelligence interacts with the development of complex social groups. We suggest that something like our postulated information integration mechanism is a product of the evolution of predator-prey survival skills in social animals, and that it is automatically activated whenever a group is co-oriented to any object, event, or issue.

Acknowledgments

This work is supported by a U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-15-1-0577. We thank Laura Halcomb for the administration of the experiments.

Author contributions

N.F. designed the experiments, analyzed the data, and took the lead in writing the paper in ongoing substantial interactions with F.B. and A.P.

Competing interests

We declare no competing interests.

Data and materials availability

All data used in the analysis will be made available upon request.

Human subjects

Our investigation involves data collected on groups of human subjects. For each subject, an informed consent was obtained after the features of the experiment's protocol were described. The investigation was approved by the IRB Committee of the University of California, Santa Barbara, and the U. S. Army DOD HRPO Official.

References

J.A. Bargh. The automaticity of everyday life. Advances in Social Cognition, 10:1–61, 1997.

J.A. Bargh and M.J. Ferguson. Beyond behaviorism: on the automaticity of higher mental processes. Psychological Bulletin, 126(6):925, 2000.

J.A. Bargh and E.L. Williams. The automaticity of social life. Current Directions in Psychological Science, 15(1):1–4, 2006.

H. Blumer. Symbolic Interactionism: Perspective and Method. University of California Press, 1969.

R. I. M. Dunbar. The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 6(5):178–190, 1998.

R. I. M. Dunbar. The social brain hypothesis. Foundations in Social Neuroscience, 5(71):69, 2002.

R. I. M. Dunbar. The social brain hypothesis and its relevance to social psychology. Evolution and the Social Mind: Evolutionary Psychology and Social Cognition, pages 21–31, 2007.

R. I. M. Dunbar. The social brain hypothesis and its implications for social evolution. Annals of Human Biology, 36(5):562–572, 2009.

R. I. M. Dunbar and S. Shultz. Evolution in the social brain. Science, 317(5843):1344–1347, 2007.

S. Epstein. Cognitive-experimental self-theory of personality. Comprehensive Handbook of Psychology: Personality and Social Psychology, 5:159–184, 2003.

E. B. Falk and D. S. Bassett. Brain and social networks: fundamental building blocks of human experience. Trends in Cognitive Sciences, 21(9):674–690, 2017.

N. E. Friedkin and E. C. Johnsen. Social Influence Network Theory: A Sociological Examination of Small Group Dynamics. Cambridge University Press, 2011.

D. Kahneman. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist, 58(9):697–720, 2003.

D. Kahneman. Thinking, Fast and Slow. Macmillan, 2011.

D. Kahneman and I. Ritov. Determinants of stated willingness to pay for public goods: A study in the headline method. Journal of Risk and Uncertainty, 9(1):5–37, 1994.

D. Kahneman, S. P. Slovic, P. Slovic, and A. Tversky. Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, 1982.

D. Kahneman, I. Ritov, and D. Schkade. Economic preferences or attitude expressions? an analysis of dollar responses to public issues. Journal of Risk and Uncertainty, 19(1):220–242, 1999.

R. Kanai, B. Bahrami, R. Roylance, and G. Rees. Online social network size is reflected in human brain structure. Proceedings of the Royal Society B: Biological Sciences, 279(1732):1327–1334, 2011.

J. E. LeDoux. Emotion circuits in the brain. Annual Review of Neuroscience, 23(1):155–184, 2000.

A. Leiserowitz. Climate change risk perception and policy preferences: The role of affect, imagery, and values. Climatic Change, 77(1-2):45–72, 2006.

C. E. Osgood, G. J. Suchi, and P. H. Tannenbaum. The Measurement of Meaning. University of Illinois Press, 1957.

C. E. Osgood, W. H. May, and M. Miron. Cross-Cultural Universials of Affective Meaning. University of Illinois Press, 1975.

S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin. Novel multidimensional models of opinion dynamics in social networks. IEEE Transactions on Automatic Control, 62:2270–2285, 2017.

J. Sallet, R. B. Mars, M. P. Noonan, J. L. Andersson, J. X. Oreilly, S. Jbabdi, P. L. Croxson, M. Jenkinson, K. L. Miller, and M. F. Rushworth. Social network size affects neural circuits in macaques. Science, 334(6056):697–700, 2011.

H. A. Simon. Models of Man: Social and Rational. Wiley, 1957.

P. Slovic, M. Finucane, E. Peters, and D. G. MacGregor. Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. The Journal of Socio-Economics, 31(4): 329–342, 2002.

T. D. Wilson. Strangers to Ourselves: Discovering the Adaptive Unconscious. Harvard University Press., 2000.

R.B. Zajonc. Feeling and thinking: Preferences need no inferences. American Psychologist, 35(2): 151, 1980.

R.B. Zajonc. Emotions. McGraw-Hill, 1998.