https://doi.org/10.1038/s44260-024-00022-y

Experimental evidence confirms that triadic social balance can be achieved through dyadic interactions

Mirta Galesic[1,2,3] ✉, Henrik Olsson[1,2] ✉, Tuan Minh Pham[1,4], Johannes Sorger[1] & Stefan Thurner[1,2,5]

Balanced triadic relationships in social groups, such that friends of friends are considered friends, are at the heart of stable human societies. Computational models of the origins of social balance typically assume that people attend to the indirect relationships between their direct social contacts. This assumption may be of limited plausability but there have been no experimental comparisons of models using different assumptions. We compare one model that assumes that people pay attention only to their direct social relationships[1], and another that assumes they try to minimize imbalance in their triadic relationships[2]. In a longitudinal group experiment with 480 interacting participants, we find that triadic social balance can be achieved even if people pay attention only to their dyadic relationships. Such empirical studies are essential for discerning between the many existing models of social dynamics and identifying the most promising pathways for further theoretical development.

To foster successful collective action, societies must achieve and maintain stable social networks. With the rise of new technologies that foster both formation and dissolution of social relationships at an unprecedented scale, it is important to understand the underlying network dynamics leading to more or less stable societies. An influential proposal has been the structural balance theory[3,4], whereby social stability can be achieved by establishing balanced cycles of relationships. The most studied cycles have been "triangles", or the relationships between any three members of a social network. A triangle is considered to be balanced if all relationships are positive (+++) or if only one relationship is positive (+--), while the remaining configurations (++- and ---) are considered to be imbalanced. This classification can be extended further by considering edge directions and incomplete triangles[5,6] as well as cycles with more than three nodes[3,7], but here we focus on the four classic types of fully connected triangles in undirected graphs.

The central hypothesis in structural balance theory is that imbalanced social cycles, such as imbalanced triangles, cause social dissonance, a term used here to align with the broader literature on dissonance reduction. This involves changing beliefs and/or social relationships to alleviate discomfort arising from inconsistencies between them[8,9]. While the term "social stress" is commonly employed in studies of structural balance, "social dissonance" better captures this broader context as highlighted in prior research. Consequently, these imbalanced cycles tend to shift toward balanced states over time[4]. As a result, the proportion of balanced triangles should increase, and the proportion of imbalanced triangles decrease over time until a steady state is reached where balanced triangles dominate in the social network. Variants of this idea have been applied in different fields, including sociology[10,11], psychology[12,13], and political science[14] to understand the dynamics of social relationships. In recent years, advances in network science and computational methods have enabled researchers to explore the structural balance theory in greater depth and detail. This has led to a renewed interest in quantitative modeling of how balance and imbalance in social networks develop over time[1,2,15,16], and how they affect social phenomena such as cooperation[17], conflict[18], and polarization[19].

A number of quantitative models have been proposed to explain changes in structural balance. Earlier models explained updating of links in social networks solely by minimizing social dissonance due to imbalanced relationships (e.g., refs. 20,21). Recently developed models acknowledge other factors contributing to changes in social balance, in particular the dynamics based on homophily of one or more opinions (e.g., refs. 1,2,15,16,22,23). These models, and other network dynamic models such as stochastic actor-oriented models[24], recognize that both opinions of and relationships between individuals can change, and that these two types of changes can affect each other.

Most quantitative models of structural balance dynamics assume that people have direct access to and take into account detailed *triadic* relationships of their nearest neighbors or even in the entire social network (but see refs. 19,25,26, discussed later). For example, the model of Pham et al.[2] assumes that people update relationships with their contacts not only based on their own agreement with these contacts but also based on how their contacts agree with each other.

[1]Complexity Science Hub, Vienna, Austria. [2]Santa Fe Institute, Santa Fe, NM, USA. [3]Vermont Complex Systems Institute, University of Vermont, Burlington, VT, USA. [4]Institute of Physics, University of Amsterdam, Amsterdam, Netherlands. [5]Section for Science of Complex Systems, Medical University of Vienna, Vienna, Austria. ✉e-mail: galesic@csh.ac.at; olsson@csh.ac.at

However, people often do not have information about agreement between their contacts in most social networks they are a part of, except perhaps in their closest social circles. Even if the information were available, for most realistic network sizes it seems implausible that people can recall and use that amount of information when deciding whether to update their relationships[27,28]. For example, a person having only 10 contacts would have to keep track and account for 45 links between these contacts. While people sometimes seek and use this information (e.g., when planning seating arrangements at an important dinner, or resolving a complicated political issue), in most everyday situations people have to rely on the information that is available and can be easily processed.

Recently, Pham et al.[1] proposed a model of social balance dynamics where the updating is based only on indivduals' *dyadic* relationships. Unlike the triadic models such as the above-mentioned Pham et al.[2], in this dyadic model people are assumed to consider only their own agreement with their contacts, and not the agreement between their contacts. Surprisingly, this model can still account for the emergence of *triadic* balance, and can reproduce empirical observations of the distribution of triangle types and the time course of triangle formation in an online multiplayer game[29]. Moreover, the model can predict tipping points at which a society becomes fragmented.

Despite the growing number of models of social balance dynamics, these models have not been compared against each other in empirical terms. Without model comparisons, ideally in longitudinal experiments that would allow for causal inferences about the underlying dynamics, it is difficult to determine which theoretical directions are promising as actual explanations of human social behavior. The models by Pham et al.[2] and Pham et al.[1] are ideal for such a comparison because they are structurally similar but have different assumptions about human cognition and give different predictions about the proportion of different triangles.

In what follows, we first review the extant experimental and longitudinal observational studies of social balance. As we will see, the experimental studies have so far been one-shot, making it difficult to investigate the time course of balance, while the longitudinal studies have been observational, making it difficult to assess causality. We then describe the two models by Pham et al. that we will compare in a longitudinal experimental study.

Experimental studies of balance largely adopt Heider's[4] POX triad conventions, where P is a focal person, O another person and X an object. X can also be another person and this triad is sometimes denoted PO_1O_2. Originally, Heider[4,30] defined eight triads, four balanced and four unbalanced. The experimental studies largely base their investigations on one of two of Heider's suggestions[12,31]. First, he suggested that "imbalance will produce tension" ([4], p 108), and second, "if no balanced state exists, then forces towards this state will arise" ([4], pp. 107–108) and balanced triads will be "preferred" ([30], p. 204). The first suggestion of "tension" has been investigated using pleasantness ratings, while the second suggestion of "balance preference" has been investigated in two ways: by eliciting expectations regarding relationships within triads, and by memory experiments where participants first learn negative or positive relationships and are later tested for their memory of them.

Experiments in the "tension" vein typically involve asking for pleasantness ratings of different triads[32–36]. Experimental manipulations often include variations in the liking between P, O and X (positive or negative; e.g., ref. 36) and the presence or absence of agreement by P and O concerning X (e.g., ref. 37). These experimental studies show evidence of one or more of three response patterns or biases. The first is the positivity bias[35,37], where triads with positive relationships are rated as more pleasant. The second is the agreement bias[35,37], where triads that have agreement between P and O concerning X are rated as more pleasant than those containing disagreement irrespective of the balance of the triad. For example, a triad that is balanced and consists of P+O, O-X, and P-X is rated as more pleasant than another balanced triad that consists of P-O, O+X, and P-X. Finally, in the balance bias[32] balanced triads are rated as more pleasant than unbalanced triads. While the empirical evidence supports each of these biases, it is unclear when one bias dominates over the other[12]. As such, these experiments do not support or rule out specific models of how people update links and opinions in relationships that we investigate here, that is if they update their relationship based on dyadic or triadic dissonance reduction. They do, however, point to the importance of agreement or homophily in link- and opinion updating, which is a crucial ingredient in the models for determining the dissonance.

In the "balance preference" vein, the predicted preference for balanced triads has been investigated by gauging participants' expectations about relationships within triadic structures. These experiments typically ask how participants would like to see relationships change for balanced and unbalanced triangles[36,38,39] or predict the values of missing relationships[40–42]. Typical experimental manipulations include, as with the pleasantness studies reviewed above, variations in the liking between P, O and X (positive or negative; ref. 38) and the presence or absence of agreement between P and O concerning X (e.g., ref. 36). As with the pleasantness ratings, there is evidence in these experiments for positivity bias[36] and agreement bias[38], but less so than for pleasantness ratings. Generally, these results point more to a balance bias than the results from the pleasantness ratings[12,43]. Again, these results do not directly speak to the potential use of dyadic or triadic dissonance in the updating of relationships. For example, people can still expect that a missing link will make a triad balanced, while at the same time update their relationships based on dyadic dissonance reduction.

Predicted preferences for balanced triads have also been studied by investigating how participants' memory errors reflect their expectations of relationships. These experiments rely on a paired-associates method, where participants are asked to learn a set of either balanced, unbalanced, or a mixed set of triads. At each trial, participants only learn one relationship[13,31,44]. The interpretation of the results from a balance perspective has been contentious with failures to replicate results and questions about how relevant paired-associate learning is for assessing if people prefer balanced triads (see reviews in Mover[12,43]). Even with these caveats, the results show a positivity bias, with positive relationships being easier to learn[31,44], some agreement bias[44], and mixed results for balance bias[13,44]. The mixed results for the balance bias might stem from the complexity of the experimental setting, with less complex experiments with fewer structures leading to more balance bias than more complex experiments that mixed many structures[43]. As with the results from the experiments reviewed above, these results do not directly speak to the potential use of dyadic or triadic dissonance in the updating of relationships.

The experimental studies reviewed in this section are mostly from the 1960s and 70s. This is the period when the bulk of the experimental studies on balance was conducted. There are a few newer studies that replicate classical findings but do not have balance as the main focus. For example, von Hecker et al.[45] replicated the classical finding of a positivity bias in a paired-associate experiment.

Investigations of the time course of balance in groups of people have so far been only observational. The data comes in the form of longitudinal surveys, databases of political relations over time, studies of online communities, and studies of organizational networks.

A classic survey data set used to investigate balance over time has been described by Newcomb[46] and includes members of a fraternity that ranked each other by attraction at weekly intervals. Using this data, Doreian and Krackhardt[47] found mixed support for the hypothesis that the proportion of balanced triangles will increase over time. While the proportion of balanced triangles (in particular +--) increased relative to the proportion of unbalanced triangles, the proportion of unbalanced --- triangles also kept increasing in this data set. Rawlings and Friedkin[48] investigated positive and negative relationships in a longitudinal survey of small communities in major U.S. cities. They investigated rules of friendship formation formulated by Rapoport[49] and found support for several rules leading to social balance, including: "a friend of a friend is a friend (+++)", "a friend of an enemy is an enemy (-+-)", and "an enemy of a friend is an enemy (+--)". Rambaran et al.[50] found support for similar rules in a survey of network formation among U.S. adolescents. In a longitudinal study of university students which

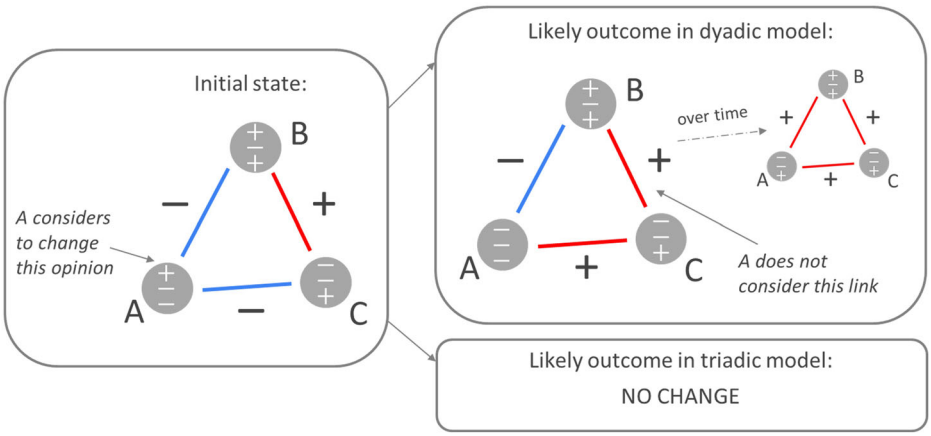

Fig. 1 | Illustration of the mechanisms underlying the link updating process in the dyadic[1] and triadic[2] models. Nodes A, B, and C are individuals. Pluses and minuses within the nodes indicate the individuals' binary opinions (positive or negative) about 3 issues. Blue lines indicate positive links and red lines indicate negative links between the individuals. The initial state (left panel) is a balanced triangle +--. What would happen if A changes the first of her opinions from + to −? Her similarity to B would decrease, justifying the already present negative link between them, but she would become more similar to person C on 2 of 3 issues, prompting her to change her link to C from − to +. Because B and C have a positive link between them, this would create an imbalanced triangle ++-. However, according to the dyadic model (top right) A does not notice this imbalance because she does not consider the link between B and C, only her own dyadic relationships with B and C. Therefore, she is likely to change her belief and the link with C. Over time, with just one more change in her belief (e.g., switching the last of her three beliefs from − to +), prompting the change of the relationship with B from − to + as well, this triangle can become a balanced +++ triangle. In contrast, in the triadic model (bottom right panel) A notices that her opinion change would produce an imbalanced triangle and does not change her belief or her link with B. As a result, the triangle remains +--. For detailed description of the mechanisms assumed by each model, please see Eqs. (1)–(4) and the accompanying text in the *Methods* section.

measured their opinions and communication networks, Linczuk et al.[23] found support for triadic influence but only when interactions were measured over multidimensional vectors of opinions, rather than for individual opinions.

Like studies of balance based on survey data, investigating balance in political relations also has a long history, starting with the illustrative examples in Harary[51]. In a longitudinal study of co-sponsorship in the U.S. Congress from 1973 to 2016, Neal[52] investigated the proportion of balanced triangles over time (+++ and +--). Overall, in both the House of Representatives and the Senate the proportion +-- was higher than for +++, and the proportion of +-- has been increasing over time while the proportion of +++ has remained constant. Most of the +++ triangles occurred within parties, while nearly all +-- triangles occurred between parties. Doreian and Mrvar[53] used data from the Correlates of War database from 1946 to 1999 to investigate networks based on alliances and conflicts. The frequency of the balanced triangle +++ was the highest through all years and it increased over time, while the other balanced triangle +-- together with the two unbalanced triangles ++- and --- had much lower absolute numbers and barely increased over time. However, Doreian and Mrvar[53] also reported the overall proportion of balanced triads, which showed both increases and decreases of balanced triads over time.

In recent years, online communities have brought new data sets that can be used to investigate balance. One important source of data are multiplayer games. In these games a diverse set of social interactions can be investigated, ranging from establishing friendships and economic relationships to forming groups, alliances, and even engaging in battles and wars. For example[29,54], studied the structural balance in an online game with more than 300,000 people. From this data it is possible to extract friendship and enmity relations and determine the level of social balance. The balanced triangles +++ and in particular +-- were overrepresented and the unbalanced triangle ++- was underrepresented with respect to a random null model. Moreover, the results showed that over time the incomplete ++ triangles close preferentially with a positive link (and become +++ triangles), and incomplete +- triangles close preferentially with a negative link (and become +-- triangles). In another online multiplayer game, Belaza et al.[14,55] found strong temporal fluctuations in the probabilities of different triangles, with the most common triangle over time being +--. Other online communities have also been studied, including product review site Epinions, Wikipedia administrator elections, and Bitcoin trust networks[56,57]. In those communities, +++ triangles are more frequent than +-- triangles[57], which are in turn less than or as frequent as ++- triangles.

In the realm of organizational networks, Askarisichani et al.[58] have recently investigated triangle balance of a network estimated from pairwise communication intensities among traders in a trading firm. They found that these networks have a tendency towards increasing balance over time, with transitions occurring from unbalanced triads to balanced triads but not the other way around.

Most observational studies find that social balance in closed social networks increases over time, but the results have been mixed regarding the proportions of different types of balanced triangles. Some studies find that ++ triangles are more prevalent than +-- triangles[52] (within parties in U.S. Congress[55], and[53], for international relationships[57], for online communities), while others find that +-- triangles outnumber +++ triangles[52] (between parties in U.S. Congress[29,55] for online multiplayer games).

In sum, the experimental studies of balance are not applicable to comparisons of the models we are considering here. The observational studies give some credence to the idea that balance should increase over time. However, the studies cover a wide range of contents, time scales, and methodologies. For example, some studies investigate fixed relationships with little or no increase in network size (e.g., the data in ref. 46), while others investigate growing networks where individual nodes can enter and exit at any time (e.g., refs. 14,29,55 and the international relations data in ref. 53), all with different ways of defining and measuring positive or negative links.

To address the lack of empirical tests of models of social balance dynamics, we conduct a longitudinal group experiment including 480 individuals and compare how well two recent models[1,2] explain the results. We first derive predictions across different parameter values in simulations with the model presented in Pham et al.[2], henceforth the *triadic model*, and the model presented in Pham et al.[1], henceforth called the *dyadic model*[1] (see details in Methods).

The two models are ideal for empirical comparisons because they are structurally similar, with both using the same statistical physics framework with the same number of parameters (two) and assuming that people are motivated to reduce dissonance in their relationships. Figure 1 illustrates the basics of each model.

Both models employ a similar belief update process, but the triadic model assumes that people try to minimize the imbalance in their triadic relationships[2] whereas the dyadic model assumes this only for their dyadic relationships[1]. Both models assume that an individual i has binary opinions s_i about G different issues ($s_i = 1$ means that i agrees with an issue, $s_i = -1$ means that i opposes it) and that they interact with K other members of their group of size N. Based on these interactions, they update their opinions as well as their relationships J_{ij} to the other members. These relations are $J_{ij} = 1$ or $J_{ij} = -1$, depending on whether i and j are friends or enemies, respectively.

The likelihood of updating is proportional to the dissonance stemming from the inconsistencies between their own and others' opinions. In the triadic model this dissonance includes inconsistencies in one or more (Q) triadic relationships of each individual. Specifically, Q is the number of triads that can change at any given update. These triads represent the social relationships that one would like to keep balanced and that can change after one's opinion changes. The other relationships that do not belong to these chosen triads are not considered for updating and hence do not change following one's opinion change.

In the dyadic model, the dissonance arises solely from dyadic relationships. It is calculated separately for one's friendly and unfriendly contacts, and the two parts are integrated after applying a homophily weight α to friendly contacts. This free parameter can be interpreted as the relative attention to friendly compared to unfriendly contacts. It represents the importance one assigns to balanced relationships within their friends' group, compared to the balance one aspires for in relationships with those who are not friends.

In both models, the dissonance affects opinion and relationship updating to the extent that people pay attention to it, as modeled by the free parameter β that is equivalent to the inverse temperature in statistical physics. This parameter is designed to capture one's overall attention to the social dissonance in a particular context. The more attention one pays to her social dissonance, the more likely is that one will update her opinions and relationships in a way that lowers the dissonance.

The group experiment enables us to compare the predictions of the two models to actual human behavior. If the dyadic model's predictions are correct, groups should be able to achieve similar levels of social balance and similar triangle proportions with or without triangle information. In the experiment, participants were arranged in 40 groups of 12 and asked to choose from among different cars to buy and companies to invest in, over four study waves (see Methods for details). Each participant made choices in eight tasks, involving a different number of cars/companies ($G = 3$ or $G = 9$ issues). For each task, participants were placed in different smaller or larger teams (with $K = 4$ or $K = 8$ members of their group). We have chosen these group sizes and values of G and K based on model simulations and previous results described in ref. 2, which have shown that these values produce noticeable differences in predicted patterns of social balance. At the same time, these values were feasible to implement in an experiment with human participants. Finally, we recruited groups of 12 participants because these groups were sufficiently large to enable comparisons of networks with degrees $K = 4$ and $K = 8$, while at the same time sufficiently small to be feasible for an experiment. Based on preliminary simulations, we decided to present 20 groups without the relevant triangle information (10 without any information and 10 with irrelevant information) and 20 groups with the relevant triangle information.

From the second study wave and on, participants were shown how their choices compared to those of their other team members in the previous wave. Critically, some groups of participants were shown only information about their dyadic relationships (whether they were friendly or unfriendly with each of their contacts, based on their past choices), while other groups were in addition shown information about their triadic relationships (whether the relationship between two of their social contacts was friendly or unfriendly). We tracked the participants' choices and relationships over time and examined the differences in the social balance and the proportion of different triangles achieved by the groups in which participants did vs. did not get triangle information.

Results

We now present the results of the simulations of the triadic and dyadic models and then compare them with the results of the longitudinal group experiment. In short, empirical evidence suggests that triadic balance can indeed be achieved only through dyadic updating, in line with the prediction of the dyadic model. Human groups achieved high level of social balance (see Eq. (5) in Methods) in just a few rounds of interactions. The proportions of different triangles (the two unbalanced types, --- and ++-, and the two balanced types, +-- and +++) were in line with the predictions of the dyadic model. We also find evidence that the critical parameters in the dyadic model, homophily weight α and attention to dissonance β, are psychologically plausible, and together can predict apparently conflicting results in the literature.

Simulations

We first compare the predictions of the two models for the level of social balance and proportion of different triangles achieved after four rounds of updates (equivalent to the four study waves in our experiment). The models can produce predictions for many different combinations of their two parameters: the proportion of triangles considered for updating (q, for the triadic model), the homophily weight (α, for the dyadic model), and the attention to dissonance (β, for both models). In addition, all predictions can be calculated for different number of issues (G) and team sizes (K). Figures S4 and S5 show predictions for a wide range of plausible combinations of these parameters.

Figure 2 shows the predictions for the parameter values that are most comparable to our group experiment. For the triadic model, the relevant parameter value is $q = 0.01$, which denotes updating of just one triangle, as it was in our experiment. For the dyadic model, the relevant parameter value is α=0.9, denoting a relatively strong homophily weight or focus on friends compared to unfriends. This corresponds to empirical observations in our experiment (see section *How psychologically plausible are model parameters*), and to observations within Szell et al.'s[54] online multiplayer game (see ref. 1). Figure 2 shows predictions for different levels of attention to dissonance (β), with values of β from 1 to 2 (denoted by green shadings) being the most realistic for our experimental setting. These values correspond to the moderate-to-high levels of attention to dissonance that we have observed empirically among our participants.

Top panels of Fig. 2 show the social balance and the proportion of different triangles predicted by the triadic model at $Q = 1$ (here represented as q=0.01 times the number of triangles, which is 10 for $K = 4$ and 75 for $K = 8$; see caption of Fig. 2), equivalent to one triangle being considered and mimicking the experiment where only one triangle was shown to participants. At the most plausible values of β (1–2, as described above), the model predicts that groups will, on average across different combinations of G and K, achieve moderately high (for $G = 9$) to high (for $G = 3$) levels of balance. The model further predicts that the proportion of +-- triangles will be larger than the proportion of +++ triangles. These patterns of results are quite similar at all levels of proportions of updated triangles q and inverse temperatures β>0, as shown in Fig. S4 (Supplementary Information, Section 1).

Bottom panels of Fig. 2 show the predictions of the dyadic model at α=.9, corresponding to empirical observations as described above. At the most plausible values of β, the model predicts that groups will, on average across different combinations of G and K, achieve moderate (for $G = 9$) to high (for $G = 3$) levels of balance. While this prediction is similar to that of the triadic model, the dyadic model's predictions for the proportion of different triangles are quite different. Instead of predicting that +-- triangles outnumber +++ triangles, the dyadic model predicts that +++ triangles will be more prevalent.

As shown in Fig. S5 in Supplementary Information, the predictions of the dyadic model for the proportion of different triangles are more nuanced than those of the triadic model. Like for the triadic model, initially predicted order of different triangles by their proportion is +--, ++-, +++, and ---. However, while for the triadic model this order stays roughly the same independently of parameters q and β, the dyadic model can produce different distributions of different triangles after a few rounds of updating,

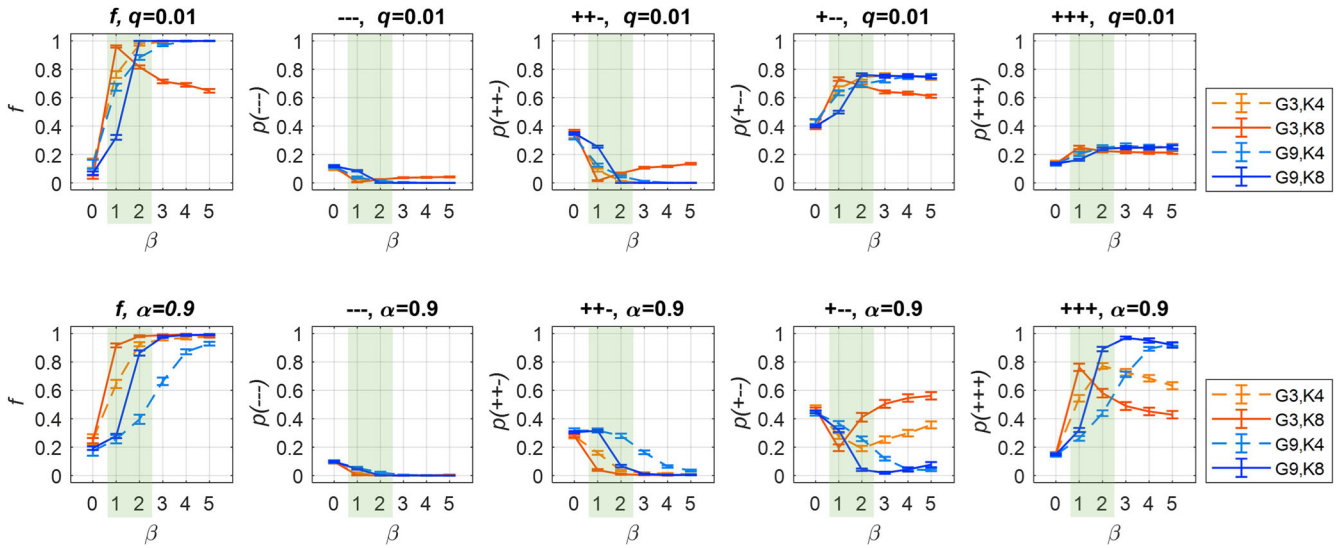

Fig. 2 | Predictions of triadic (top) and dyadic (bottom) models for the level of balance (f) and proportions of different triangles after four updating rounds. Shown are predictions for different number of issues ($G = 3$, light red or $G = 9$, light blue), and team sizes ($K = 4$ or $K = 8$), equivalent to the conditions in the group experiment. Top: Predictions of the triadic model assuming that each individual updates only one of their triangletime step, like in the group experiment s in each ($Q = 1$ or more generally $q = \lceil Q/N^T \rceil$, where N^T is the total number of triangles in a group; see Figure S4 for more details). Bottom: Predictions of the dyadic model, assuming high levels of attention to friendly vs. unfriendly team members α, as observed in Szell et al.[29]. All predictions are shown for different levels of attention to dissonance β. Green shaded areas denote moderate-to-high levels of attention that were observed in our experiment. Error bars denote the 95% confidence intervals.

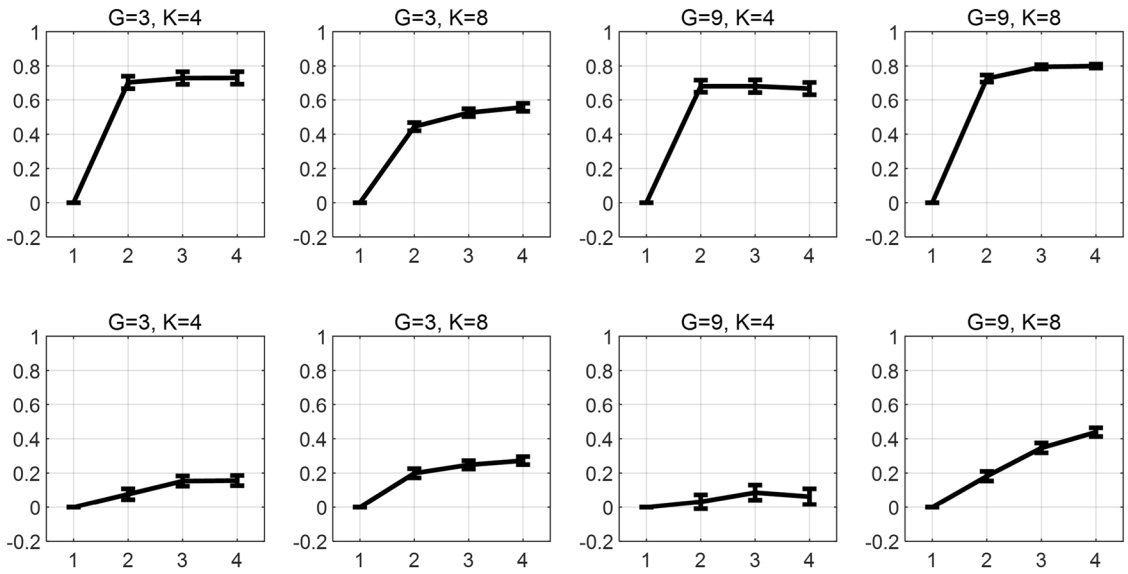

Fig. 3 | Normalized balance ($\hat{f}$) over time predicted by the triadic (top) and dyadic (bottom) model for situations in which participants decide about different number of issues G, and are embedded in teams of different sizes =K. Error bars represent confidence intervals. The initial balance f is set to $\hat{f} = 0$ for easier comparison of tasks and experimental conditions.

depending on the level of homophily weight (α). At lower levels of α the model predicts higher or equal proportion of +-- triangles compared to +++ triangles. However, with sufficiently high α, its prediction flips towards a higher proportion of +++ than +-- triangles, as observed in Fig. 2.

We next compare the predictions of the two models for the level of social balance and proportion of different triangles over each of the four time steps, corresponding to the four study waves in our experiment. We assume the parameter values that most correspond to our experimental setting, as described above (for the triadic model, $q = 0.01$; for the dyadic model. α=0.9; for both $\beta = 0.2$). Both models predict an increase in balance over time, but the dyadic model predicts a slower increase than the triadic model (Fig. 3). When it comes to the proportion of different triangles (Fig. 4), the predictions of the triadic model are very different than those of the dyadic model, with +-- triangle dominating all waves in the triadic model, and +++ triangles prevailing in the dyadic model predictions.

Group experiment

Next, we use the data from our longitudinal group experiment to investigate whether people use triangle information to achieve social balance, how well the two models predict the final balance and proportions of different triangles, and how psychologically plausible model parameters are.

Do people need triangle information to achieve balanced triangles?.

We first examined the differences between groups of participants who received either no information about triangles, irrelevant information, or relevant information about one of the triangles they were a part of.

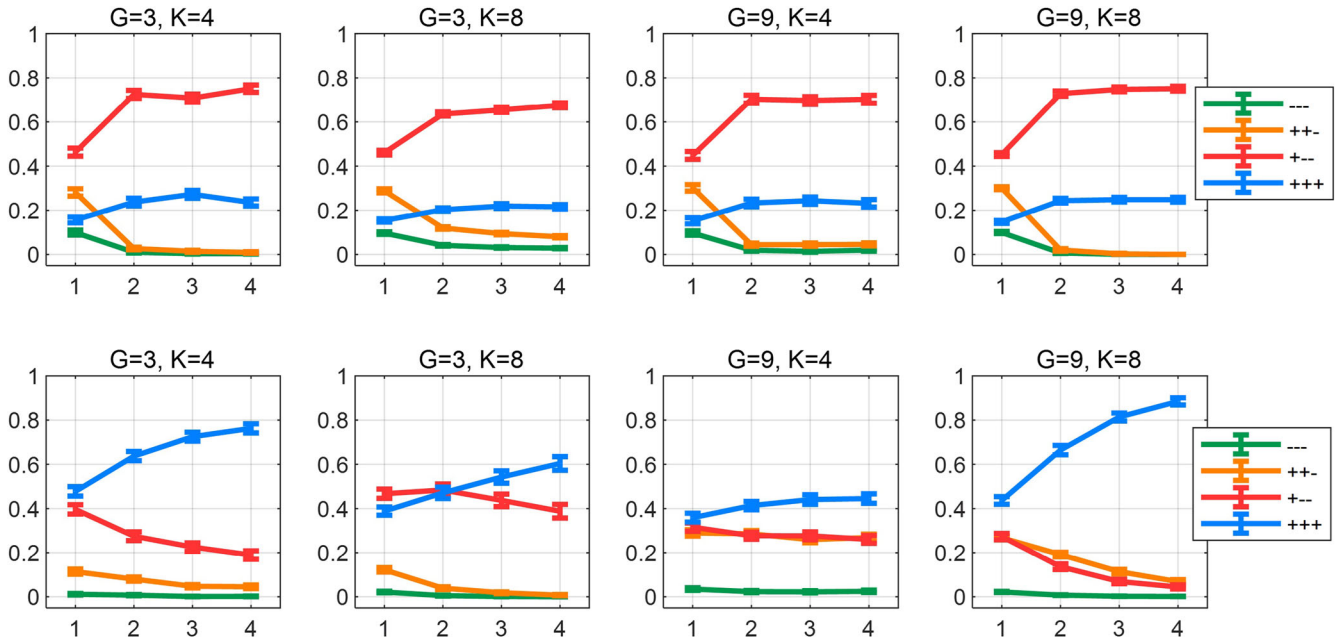

Fig. 4 | Proportion of different triangles over time predicted by the triadic (top) and dyadic (bottom) model for situations in which participants decide about different number of issues *G*, and are embedded in teams of different sizes *K*. The differently colored lines represent the triangles --- (green), ++- (orange), +-- (red), +++ (blue). Error bars represent confidence intervals.

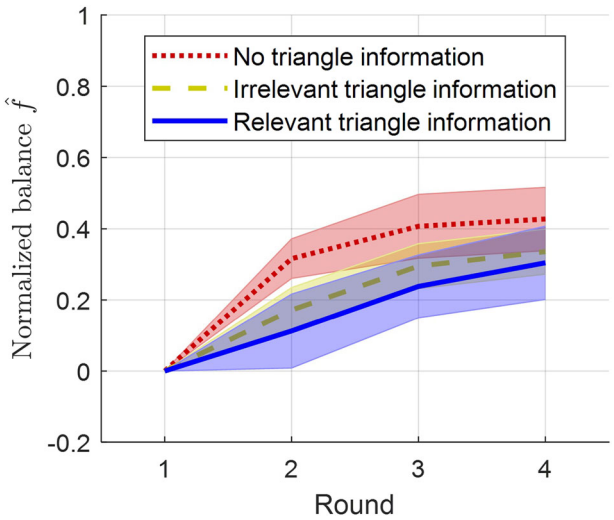

Fig. 5 | Effect of information about triangle relationships on the normalized balance ($\hat{f}$) achieved by groups of participants receiving no information (red), irrelevant information (yellow), or relevant information about one of their triangles (blue). Shaded areas represent the 95% confidence intervals. The initial balance (on average acros s groups, $f_1 = 0.27$) is set to $\hat{f}_1 = 0$ for easier comparison of tasks and experimental conditions. Note that balance forms very rapidly, reaching $f_4 = 0.62$ ($\hat{f}_4 = 0.35$ after normalization) in only three study waves. The achieved level of balance corresponds better to the predictions of the dyadic model than of the triadic model (Fig. 3).

Participants in all three conditions were affected by the information they received about their groups (Table S1), but more so if they received either irrelevant or relevant triangle information. However, as shown in Fig. 5, the mean trends in balance were not systematically different between these different experimental conditions, with confidence intervals largely overlapping. This was so independently of the number of issues *G*, the number of contacts *K*, or the topic of the task (cars or companies; see Fig. S6). If anything, the groups without any triangle information achieved higher levels of social balance faster than the groups with either relevant or irrelevant triangle information. At the same time, the balance has increased from the first to the last round of the experiment, in all three conditions.

The proportion of balanced triangles +++ increased throughout the experiment in all three experimental conditions (Fig. 6), while the proportion of unbalanced triangles as well as +-- triangles decreased. The empirical patterns in Fig. 6 clearly correspond better to the predictions of the dyadic model (Fig. 4), both in terms of the achieved levels of balance, and in terms of the proportions of different triangles produced over time. The same pattern occurs in experimental conditions with different number of issues (*G*), the number of contacts (*K*), and with tasks involving cars or companies (Fig. S7). More generally, participants gained more friendly than unfriendly contacts over successive rounds of the experiment (Fig. S8).

How well are the final balance and triangle statistics predicted by models of social balance?. Next, we compare the empirical results in the last study wave, for the 20 groups who received relevant triangle information, with the model predicted patterns. The empirical results are summarized as boxplots in Fig. 7. The level of balance across different number of issues (*G*) and number of contacts (*K*) is around 0.6, with the +++ triangles being by far most prevalent (60% of triangles), followed by about 20% of both balanced +-- or unbalanced ++- triangles, and a rare occurrence of --- triangles.

To compare the experimental results with the predictions of the triadic model shown in the top panels of Fig. 2, we average the model predictions in green shaded areas of these panels, corresponding to predictions at moderate-to-high level of attention (β of 1 or 2). Comparing these average predictions, denoted by Xs in Fig. 7, with box plots showing experimental results, we see that this model corresponds to the level of balance *f* observed in the experiment for different values of *G* and *K* (left-most column of Fig. 7) as well as the relative proportion of unbalanced triangles (the second and third columns of Fig. 7).

However, the triadic model predictions for the relative proportion of different balanced triangles do not correspond to the empirical observations. As shown in the last two columns of Fig. 7, the model always predicts many more +-- than +++ triangles, opposite from the empirical observations. This pattern of predictions is independent of the values of β (see Fig. S4).

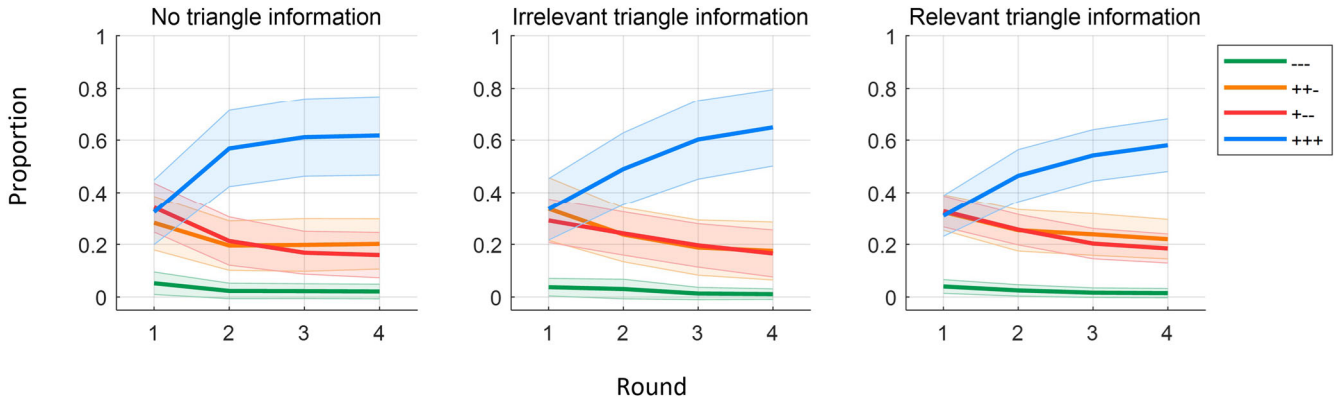

Fig. 6 | Proportion of different triangles occurring in the group experiment over time, by experimental conditions and tasks. The triangles in each wave are based on links J_{ij} that were updated based on participants' answers in that wave according to Eq. (3). The differently colored lines represent the triangles --- (green), ++- (orange), +-- (red), +++ (blue). The empirically observed patterns are similar to the predictions of the dyadic model, but not to those of the triadic model (Fig. 4).

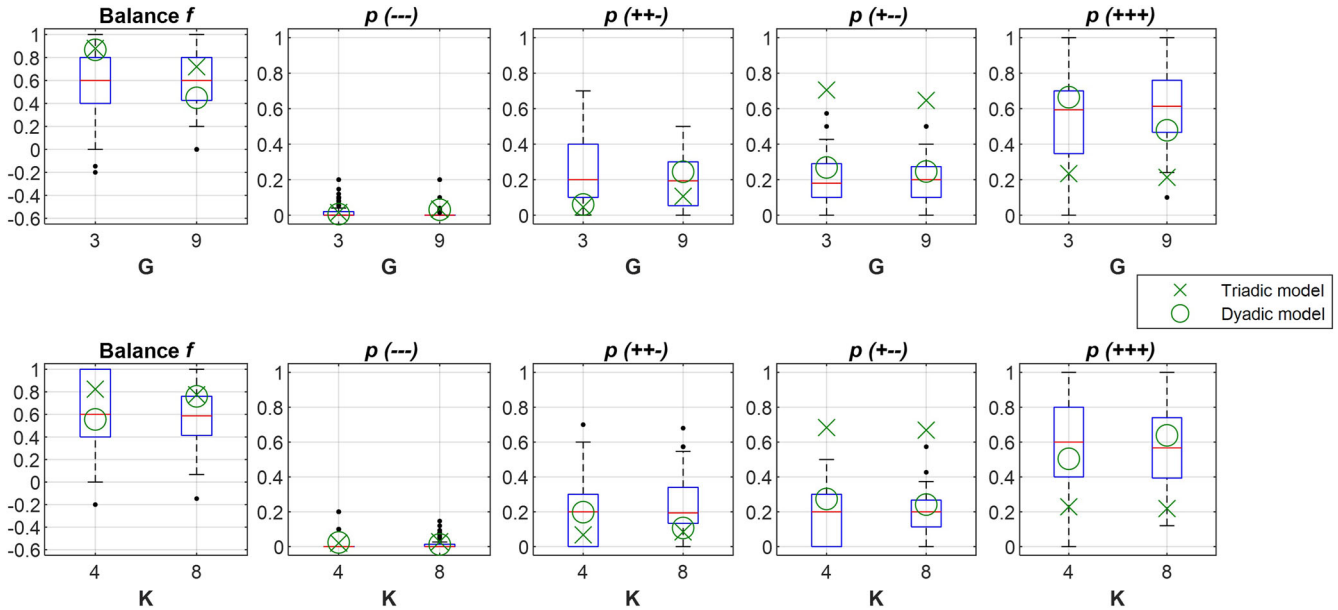

Fig. 7 | Relation of the number of issues participants had to decide about in each task (*G*, top panels) and the number of other participants they worked with (*K*, bottom panels), with the social balance (*f*, column 1) and the proportion of different triangles occurring in the group experiment (columns 2–5). For each box, the central red line is the median, the edges are the 25th and 75th percentiles, the whiskers extend to the most extreme data points within another box length on each side, and the outliers are plotted individually as dots. Green Xs denote the mean predictions of the triadic model for $q = 0.01$, while green circles denote the mean predictions of the dyadic model for α=0.9 (see Fig. 2). For both models, predictions are averaged over levels of attention $\beta \in \{1, 2\}$ (the areas of Fig. 2 denoted by green shaded areas). Shown are results for the 20 participant groups who received the relevant triangle information; patterns are similar for all 40 groups (see Fig. S9).

We can do the same comparison of the experimental results with the predictions of the dyadic model in the bottom panels of Fig. 2. The model predicts the empirically observed level of balance at β around 1 or 2 (see circles in the left-most column of Fig. 7). In addition, the model predicts the order of relative proportions of different triangles. The +++ triangles are predicted to be the most prevalent, followed by +--, ++-, and finally --- triangles, in line with the experimental results (see circles in the four right columns in Fig. 7). This is an important difference compared to the triadic model, which predicts that +-- triangles will be more prevalent than +++ triangles. One pattern predicted by the dyadic model is not reflected in our empirical data: the model predicts a higher balance for groups with larger K and smaller G, but we find no consistent patterns in our data (see Figs. 5–7).

How psychologically plausible are model parameters?. The dyadic model assumes that its parameters α and β correspond to actual psychological mechanisms underlying the dynamics of social balance. To test this, we collected participants' subjective reports of their level of attention to friendly vs. unfriendly team members (α^*) and their perceived level of distraction and difficulty (β^*, see *Methods* for more details). In line with our findings that the model is best aligned with the experimental data when α and β are moderately high, most participants reported paying equal (64%) or more (28%) attention to friendly than to unfriendly contacts (2%), and most (more than 90%) reported that they were able to pay attention to the task (that is, that they were not distracted and that the task was very easy).

Even though it is not possible to map the absolute values of participants' subjective reports directly to the parameter values, we still expect that their reports correlate in specific ways with the achieved level of balance and the proportion of different triangles. In particular, from the model predictions for different levels of α and β, shown in Figs. 2 and S5, we expect a positive relationship of the subjective estimates of these parameters, α^* and β^* with the overall social balance and with the proportion of +++ triangles.

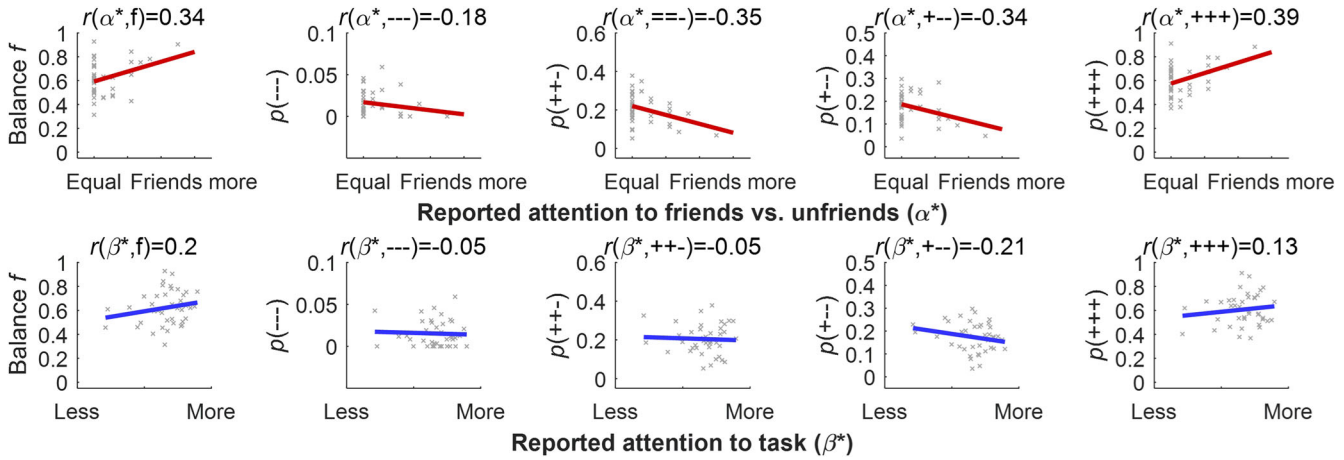

Fig. 8 | Relationship of the subjective estimates of model parameters α (α*, reported attention to friendly compared to unfriendly team members) and β (β*, inverse of reported distraction and difficulty of tasks). The estimates were measured in the last study wave and averaged for each of the 40 participant groups, with the social balance (f) and the proportion of different triangles in the last study wave. Results for all four study waves are shown in Figs. S10 and S11.

We expect a negative relationship of the subjective estimates α* and β* with the proportion of the other types of triangles.

As shown in Fig. 8, remarkably, we find the expected positive relationships of balance and proportion of +++ triangles with the participants' subjective reports of model parameters α* and β* (averaged for each of the 40 groups). Also as expected, we find small to negative relationships between these subjective estimates and other types of triangles, including +--. These results are in line with the assumptions of the dyadic model in Pham et al.[1], providing further evidence that this is a promising theoretical framework for understanding the dynamics of social balance.

Discussion

In this study, we addressed the lack of longitudinal experimental studies of balance dynamics as well as the lack of comparison of models of this dynamics. Previous experimental studies of social balance were one-shot, preventing investigations of the dynamics of balance, while the longitudinal studies were observational, making it difficult to assess causality. We compare two recent models of balance dynamics, a triadic model that assumes that people try to minimize imbalance in their triadic relationships[2], and a dyadic model that assumes people pay attention only to their direct social relationships[1].

We find support for a provocative prediction of the dyadic model by Pham et al.[1]: groups are able to achieve social balance within a few interaction rounds even without the information about triangle relationships in their social networks. This is an important finding because it shows that social stability is possible even with limited knowledge of and capacity to account for the many different indirect relationships people have. The proportion of different triangles and the overall balance achieved in our experiment (Figs. 5 and 6) are in line with the dyadic model predictions (bottom panels of Figs. 2–4), assuming plausible values of model parameters α and β in line with their subjective estimates (Section "How psychologically plausible are model parameters?") and past studies[29].

Our results stress the importance and demonstrate psychological plausibility of two critical parameters in the Pham et al.[1] model: the differential attention to friendly vs. unfriendly social contacts (homophily weight α) and the overall attention to the social dissonance in a particular context (β). As shown in Figure S5, when the attention to both friendly and unfriendly contacts is the same (α=0.5), the distribution of different triangles stays the same as it initially was, with +-- triangles outnumbering the +++ triangles. When the attention to friendly contacts is higher, as it was in our experiment, the +++ triangles become the most prevalent. Differences in this homophily weight might help explain differences in empirical findings about the proportion of different triangles. While many studies find that +++ triangles dominate, some studies find that the most common triangles are +-- (see Introduction). Investigating the empirical differences in the motivations and the ability to attend to friendly vs. unfriendly social contacts in different societal contexts would therefore be an interesting avenue for future research. In addition, our results suggest that social balance cannot be achieved unless individuals pay at least some attention to the social dissonance and update their beliefs and relationships accordingly (parameter β in the model).

Note that we have presented a qualitative comparison of predicted and empirical values rather than quantitative model fitting. Because the models were not designed to mimic specific psychological processes, but to provide sufficient conditions for achieving structural balance, conducting quantitative fits of model parameters to the data seems inappropriately precise[59]. The qualitative comparison allows us to discern which of the models is more promising for future development but at the same time does not tempt us to overgeneralize from abstract models "as if" they represented actual psychological mechanisms[60]. Both the triadic and dyadic models[1,2] (respectively) were designed on a relatively abstract level, aiming to show sufficient assumptions for achieving global structural balance on the basis of individuals minimizing cognitive dissonance, rather than to reproduce the specific psychological mechanisms underlying participants' choices. While the dyadic model is more cognitively realistic than the triadic model in that it does not assume knowledge of or attention to triadic social relationships, neither model incorporates cognitive details that likely affect actual human performance in this experiment. For example, both models assume that people calculate dissonance based on the difference between all of their opinions and all of the opinions of all of their group members. Even for a task with small G and K, this would require attending to $G = 3$ answers of each of the $K = 4$ contacts, and comparing them with each of one's own answers. This requires 12 comparisons in each study wave just for this task, in addition to other considerations (triangle relationships in Eq. (1), or who is friendly or not friendly in Eq. (2)) for each task. For large G and K, the number of comparisons rises to 72 per task, which is almost certainly not what people are doing.

Our study suggests guidelines for further theoretical development and empirical research on structural balance. While our experimental results clearly support dyadic model over the triadic one, details of an empirical setting certainly matter. They can affect the attention paid to friends, the level of attention to any dissonance and the overall distraction, the knowledge about the relationships in one's social environment, the degree to which it involves coordination or conflict, how different outcomes carry different consequences and other relevant factors that are hard to control. For example, the dyadic model predicts differences in balance achieved in conditions with different levels of G (number of items or issues to have an opinion on) and K (number of social contacts or team size), which we do not

observe empirically. A promising theoretical direction can be to adjust the assumption that people calculate social dissonance based on their and others' opinions about many different issues at hand. When the number of issues G is large, it is possible that only a few issues at a time are used to calculate the dissonance. Further refinements can include limiting the number of social contacts K that participants pay attention to, as they might be able to attend to only a few other individuals. Finally, model refinements could include stubbornness regarding own initial opinions. People might experience dissonance when their new beliefs deviate from their previous ones, and this dissonance may add on to the overall social dissonance. These additional assumptions, while plausible, should be tested empirically by measuring people's attention span for items and other group members, and their commitment to their initial beliefs.

Our results also provide an indirect support for other models of belief and social dynamics that show plausible collective patterns of beliefs emerging from dyadic social interactions in social networks, whereby two agents aim to establish a balanced relationship between themselves and their beliefs on common issues[19,25,26]. Similarly to the dyadic model, individuals in these models are assumed to adjust several of their beliefs and their dyadic relationships simultaneously, following simple homophily-based updating like here[26], or other mechanisms such as associative diffusion[25] or structural balance on the level of two individuals and their opinion vectors (Schweighofer et al.[19]). While these studies did not explore indicators of social balance such as those presented here (in particular, the proportion of different types of triangles), future experimental research could compare those models with the dyadic model.

In summary, we have shown that it is possible to empirically observe the formation of social balance in group experiments. Despite the large number of quantitative models, there is a huge need for empirical model comparisons[61,62]. Rather than developing further models without empirical tests, substantial progress in understanding human sociality can be achieved by a tight coupling of model development with studies of actual cognitive mechanisms and behavior.

Methods

Simulations

We simulated the dynamics of the triadic and dyadic models to compare their outcomes with those of the social group experiment. The simulations use the same assumptions as the group experiment used to test them (see Section "Longitudinal group experiment"). Specifically, we assumed teams of $N = 12$ individuals, with each individual connected to either $K = 4$ or $K = 8$ other individuals and having binary opinions $[-1,1]$ on $G = 3$ or $G = 9$ issues. To construct initial social networks of these teams, we iterated over different random graphs of size N and degree K until we obtained networks with the maximum number of triangles (10 triangles for $K = 4$, 75 triangles for $K = 8$; see Fig. S3 in Supplementary Information for visualizations of the network structures). Each node was involved in a similar number of triangles. Of note, we used those same networks in the group experiment described below. We ran the simulations over four time steps (again, as in the group experiment), and average them over 100 independent runs for each combination of parameters.

Both models were initialized by assigning each individual an opinion vector $\mathbf{s}_i$ composed of randomly chosen opinions (−1 or 1) on G issues, and connecting each individual with K other group members through links J_{ij} to establish the social network structure. For the triadic model, at each time step the local social dissonance function for each individual is calculated as

$$H^{(i)} = -\frac{1}{G}\sum_{j} J_{ij}\mathbf{s}_i \cdot \mathbf{s}_j - \sum_{(j,k)_{Q_i}} J_{ij}J_{jk}J_{ki}, \tag{1}$$

with Q_i denoting the number of randomly chosen triangles that an individual i is trying to balance, of all triangles that the individual has with j and k other individuals.

For the dyadic model, the local social dissonance function for each individual is

$$H^{(i)} = -\frac{\alpha}{G}\sum_{j:J_{ij}=1} \mathbf{s}_i \cdot \mathbf{s}_j + \frac{1-\alpha}{G}\sum_{j:J_{ij}=-1} \mathbf{s}_i \cdot \mathbf{s}_j, \tag{2}$$

with the first sum including all friendly contacts of i ($J_{ij} = 1$), and the second all unfriendly contacts ($J_{i\,j} = -1$), weighted by the parameter α that can be operationalized as a *homophily weight* (attention to friendly compared to unfriendly contacts).

For both models, the edge weights were initially calculated and subsequently updated according to

$$J_{ij} = \mathrm{sign}(\mathbf{s}_i \cdot \mathbf{s}_j), \tag{3}$$

For the triadic model, at each time step we (i) randomly choose one individual, flip one of their G opinions ($s_i \to -s_i$), (ii) recalculate the values of their social dissonance $H^{(i)}$ [1 or 2] for both models and (iii) update the edge weights J_{ij} according to ref. 7. The dyadic model follows the same steps except that the order of steps (ii) and (iii) is reversed, in line with its original version in[1]. This does not change the results. We then (iv) accept the opinion flip with the probability:

$$p = \min\left\{e^{-\beta\Delta H^{(i)}}, 1\right\}, \tag{4}$$

where $\Delta H^{(i)}$ is the difference between updated and previous social dissonance, and β is the inverse temperature parameter that captures the *attention to dissonance* stemming from the current constellation of opinions and relationships. If the opinion flip is rejected, the recalculated edge weights are also rejected. We repeat (i)–(iv) above until all individuals' opinions and edge weights have been updated before continuing to the next time step.

We then calculate the level of social balance achieved at the end of each simulation as the relative difference of balanced n_+ and unbalanced n_- triangles:

$$f = (n_+ - n_-)/(n_+ + n_-), \tag{5}$$

where n_+ and n_- are the numbers of balanced and unbalanced triangles, respectively. We also record the proportions of different types of triangles: the two balanced types (+++ and +--) and the two unbalanced types (++- and ---).

Longitudinal group experiment

To compare the triadic and the dyadic model described in the previous section, we collected data from human groups in a longitudinal experiment designed to be very similar to the simulation setting. We recruited 480 participants from the crowdsourcing platform Mechanical Turk using Cloud Research platform that enables guardrails for known problems with unique and truthful identities, and makes it easier to conduct longitudinal research with the same participants. Among the participants, 51% were male, 35% had less than bachelor degree, and average age was 41 years (SD = 12.3), ranging from 19 to 78 years. Participants received fixed incentives of $1.5 for participating in wave 1, $1.5 for wave 2, $2 for wave 3, and $3 for wave 4. Each wave took about 5 min to complete. There was no incentive to be more responsive to the choices of and relationships with team members, to avoid producing responses that accord with the research questions.

Participants were divided in 40 groups of 12 and have received 8 tasks involving choices of car types and industry sectors in each of 4 study waves ~10 days apart. For each task, participants were assigned to a different team composed of a random subset of their group members, and embedded in network structures that were the same as those used in simulations described above (see Fig. S3 in Supplementary Information). These teams remained the same across all study waves.

The 8 tasks (see examples in Fig. S1, and the whole questionnaire in Supplementary Information, Sections 2 and 3) were characterized by a different combination of the following three *within-subject* factors: (i) number of issues the participants had to decide about ($G = 3$ or $G = 9$ car types or industry sectors), (ii) the number of other participants in the team ($K = 4$ or $K = 8$), and (iii) topic – choosing the type of car to buy for their company (e.g., coupe or sedan), or industry sector (e.g., tech or pharmaceutical industry) they would like to invest in. The two topics – cars and companies – served as replications of the same combinations of G and K. In each wave, tasks with the particular combination of G and K contained the same types of cars and industry sectors but with different examples of cars and industries (for example, Mini Hardtop and Volkswagen Beetle for coupes, Honda Civic and Nissan Altima for sedans, Google or Microsoft for tech industries, Pfizer or Roche for pharmaceutical industries; see issues and examples for each topic in Supplementary Information, Section 3). In this way we tried to make the tasks very similar regarding content and structure across waves, but sufficiently novel each time. The order of tasks was randomized across participants and waves.

To test whether the information about triangles affected the formation of social balance, the study included a *between-subject* manipulation with three conditions, one experimental and two controls (see Fig. S2 in Supplementary Information). In the experimental condition (bottom row of Fig. S2), 20 groups of participants got information about one of the triangles they were involved in and were told (truthfully) that these connections were based on the result of their choices in the tasks in the previous wave, and represented friendly (full lines) and unfriendly (dashed lines) relationships. In the first control condition (top row of Fig. S2), 10 groups of participants got no information about their triangles at all. To make sure that differences in cognitive load do not affect the results, in the second control condition (middle row of Fig. S2), 10 groups got information about one of the triangles they were involved in but were told (again, truthfully) that this information is essentially irrelevant for the task of choosing cars and companies. Specifically, this second group was told that the other two members of the triangle were connected because they were randomly chosen to participate in a fundraising event, with full lines denoting an event for a local children's hospital, and dashed line for a local zoo.

After each study wave, we updated the participants' friendly and unfriendly links J_{ij} using Eq. (3) above. In each subsequent wave, we showed the participants their answers as well as the answers of their friendly and unfriendly contacts (see Fig. S1 for examples of this display) and tracked how much they updated their initial beliefs.

To make sure that all participants understood the way these answers and the triangle information were presented, we included detailed instructions and examples, as well as four practice tasks at the beginning of the first wave and two at the beginning of each of the subsequent waves (see Supplementary Information, Section 2). Participants had to answer 10–12 (depending on whether triangle information was present in their condition or not) test questions correctly after each of the practice tasks to be allowed to proceed to each wave of the study.

Besides tasks involving choices between cars and companies, participants were also asked additional questions designed to elicit the psychological mechanisms assumed by model parameters α in Eq. (2) and β in Eq. (4). The parameter Q_i in Eq. (1) was set to 1 for all participants who received triangle information, as they all received information about only one of their triangles. For α, participants were asked at the end of each wave to report how much attention they paid to their friendly vs. unfriendly team members ("Overall, across different tasks, how much attention did you pay to your friendly and unfriendly team members?"). We calculated the median of the responses to this question over waves 2–4 (wave 1 was excluded as the social information was not yet available) and used it as empirical parameter α^*, explored in Fig. 8. For β, we asked participants about their level of distraction at the end of each wave ("How distracted were you while answering questions?"), as well as several questions about the perceived difficulty of the tasks at the end of the study ("How easy or difficult was it to understand all the instructions?", "How easy or difficult was it to answer the questions about types of cars?", "How easy or difficult was it to answer the questions about industry sectors?", and "How easy or difficult was it to understand the information about others in your group?"). We calculated the inverse average of responses to these questions, assuming that lower distraction and better understanding of the tasks related to a higher attention to the task, or lower 'temperature' in Eq. (4). We used this average response as the empirical parameter β^*, explored in Fig. 8.

The study was conducted in June and July of 2022. All materials and procedures were extensively pretested in several smaller studies on the same population. These pretests also served as benchmark for determining the final sample sizes appropriate for discerning the patterns in different experimental conditions. The study was approved by the Institutional Review Board of the University of New Mexico (no. 1331148-4) and was reviewed positively by the Research Ethics Committee of the Vienna University of Technology.

Data availability

All data needed to evaluate the conclusions in the paper are available at https://osf.io/ur2b4/.

Code availability

All code needed to evaluate the conclusions in the paper are available at https://osf.io/ur2b4/.

Received: 30 January 2024; Accepted: 13 November 2024;
Published online: 08 January 2025

References

1. Pham, T. M., Korbel, J., Hanel, R. & Thurner, S. Empirical social triad statistics can be explained with dyadic homophylic interactions. *Proc. Natl Acad. Sci.* **119**, e2121103119 (2022).
2. Pham, T. M., Alexander, A. C., Korbel, J., Hanel, R. & Thurner, S. Balance and fragmentation in societies with homophily and social balance. *Sci. Rep.* **11**, 17188 (2021).
3. Cartwright, D. & Harary, F. Structural balance: a generalization of Heider's theory. *Psychol. Rev.* **63**, 277 (1956).
4. Heider, F. Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
5. Holland, P. W. & Leinhardt, S. A method for detecting structure in sociometric data. *Am. J. Sociol.* **76**, 492–513 (1970).
6. Holland, P. W. & Leinhardt, S. Local structure in social networks. *Sociol. Methodol.* **7**, 1–45 (1976).
7. Aref, S. & Wilson, M. C. Measuring partial balance in signed networks. *J. Complex Netw.* **6**, 566–595 (2018).
8. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford University Press, 1957).
9. Gawronski, B. Back to the future of dissonance theory: cognitive consistency as a core motive. *Soc. Cogn.* **30**, 652–668 (2012).
10. Davis, J. A. Structural balance, mechanical solidarity, and interpersonal relations. *Am. J. Sociol.* **68**, 444–462 (1963).
11. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).
12. White, C. J. M. Factors affecting balance, agreement and positivity biases in POQ and POX triads. *Eur. J. Soc. Psychol.* **9**, 129–148 (1979).
13. Zajonc, R. B. & Burnstein, E. The learning of balanced and unbalanced social structures. *J. Personal.* **33**, 153–163 (1965).
14. Belaza, A. M. et al. Statistical physics of balance theory. *PLoS ONE* **12**, e0183696 (2017).

15. Górski, P. J., Bochenina, K., Hołyst, J. A. & D'Souza, R. M. Homophily based on few attributes can impede structural balance. *Phys. Rev. Lett.* **125**, 078302 (2020).
16. Pham, T. M., Kondor, I., Hanel, R. & Thurner, S. The effect of social balance on social fragmentation. *J. R. Soc. Interface* **17**, 20200752 (2020).
17. Traag, V. A., Van Dooren, P. & De Leenheer, P. Dynamical models explaining social balance and evolution of cooperation. *PLoS ONE* **8**, e60063 (2013).
18. Lerner, J. Structural balance in signed networks: separating the probability to interact from the tendency to fight. *Soc. Netw.* **45**, 66–77 (2016).
19. Schweighofer, S., Schweitzer, F. & Garcia, D. A weighted balance model of opinion hyperpolarization. *J. Artif. Soc. Soc. Simul.* **23** (2020).
20. Antal, T., Krapivsky, P. L. & Redner, S. Dynamics of social balance on networks. *Phys. Rev. E* **72**, 036121 (2005).
21. Marvel, S. A., Kleinberg, J., Kleinberg, R. D. & Strogatz, S. H. Continuous-time model of structural balance. *Proc. Natl Acad. Sci.* **108**, 1771–1776 (2011).
22. Górski, P. J., Atkisson, C. & Holyst, J. A. A general model for how attributes can reduce polarization in social groups. *Netw. Sci*. **11**, 536–559 (2023).
23. Linczuk, J., Górski, P. J., Szymanski, B. K. & Hołyst, J. A. Multidimensional attributes expose Heider balance dynamics to measurements. *Sci. Rep.* **13**, 15568 (2023).
24. Snijders, T. A. Stochastic actor-oriented models for network dynamics. *Annu. Rev. Stat. Appl.* **4**, 343–363 (2017).
25. Goldberg, A. & Stein, S. K. Beyond social contagion: associative diffusion and the emergence of cultural variation. *Am. Sociol. Rev.* **83**, 897–932 (2018).
26. Rodriguez, N., Bollen, J. & Ahn, Y. Y. Collective dynamics of belief evolution under cognitive coherence and social conformity. *PLoS ONE* **11**, e0165910 (2016).
27. Dunbar, R. I. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **22**, 469–493 (1992).
28. Gonçalves, B., Perra, N. & Vespignani, A. Modeling users' activity on Twitter networks: validation of Dunbar's number. *PLoS ONE* **6**, e22656 (2011).
29. Szell, M., Lambiotte, R. & Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. Natl Acad. Sci.* **107**, 13636–13641 (2010).
30. Heider, F. *The Psychology of Interpersonal Relations* (Wiley, 1958).
31. Zajonc, R. B. & Sherman, S. J. Structural balance and the induction of relations. *J. Personal.* **35**, 635–650 (1967).
32. Aderman, D. Effects of anticipating future interaction on the preference for balanced states. *J. Personal. Soc. Psychol.* **11**, 214 (1969).
33. Cacioppo, J. T. & Petty, R. E. Effects of extent of thought on the pleasantness ratings of p-o-x triads: evidence for three judgmental tendencies in evaluating social situations. *J. Personal. Soc. Psychol.* **40**, 1000–1009 (1981).
34. Gerard, H. B. & Fleischer, L. Recall and pleasantness of balanced and unbalanced cognitive structures. *J. Personal. Soc. Psychol.* **7**, 332–337 (1967).
35. Gutman, G. M. & Knox, R. E. Balance, agreement, and attraction in pleasantness, tension, and consistency ratings of hypothetical social situations. *J. Personal. Soc. Psychol.* **24**, 351–357 (1972).
36. Rodrigues, A. Effects of balance, positivity, and agreement in triadic social relations. *J. Personal. Soc. Psychol.* **5**, 472 (1967).
37. Whitney, R. E. Agreement and positivity in pleasantness ratings of balanced and unbalanced social situations: a cross-cultural study. *J. Personal. Soc. Psychol.* **17**, 11–14 (1971).
38. Burnstein, E. Sources of cognitive bias in the representation of simple social structures: balance, minimal change, positivity, reciprocity, and the respondent's own attitude. *J. Personal. Soc. Psychol.* **7**, 36–48 (1967).
39. Wyer, R. S. & Lyon, J. D. A test of cognitive balance theory implications for social inference processes. *J. Pers. Soc. Psycho*. **16**, 598–618 (1970).
40. Crockett, W. H. Balance, agreement, and subjective evaluations of the POX triads. *J. Personal. Soc. Psychol.* **29**, 102–110 (1974).
41. Fuller, C. H. Comparison of two experimental paradigms as tests of Heider's balance theory. *J. Pers. Soc. Psychol.* **30**, 802–806 (1974).
42. Rodrigues, A. The biasing effect of agreement in balanced and imbalanced triads. *J Pers*. **36**, 138–153 (1968).
43. Crockett, W. H. Balance, agreement, and positivity in the cognition of small social structures. in *Advances in Experimental Social Psychology*. Vol. 15, 1–57 (Academic Press, 1982).
44. Rubin, Z. & Zajonc, R. B. Structural bias and generalization in the learning of social structures. *J. Personal.* **37**, 310–324 (1969).
45. von Hecker, U., Hahn, U. & Rollings, J. Spatial representation of coherence. *J. Exp. Psychol.: Gen.* **145**, 853–871 (2016).
46. Newcomb, T. M. *The Acquaintance Process* (Holt, Rinehart & Winston, 1961).
47. Doreian, P. & Krackhardt, D. Pre-transitive balance mechanisms for signed networks. *J. Math. Sociol.* **25**, 43–67 (2001).
48. Rawlings, C. M. & Friedkin, N. E. The structural balance theory of sentiment networks: elaboration and test. *Am. J. Sociol.* **123**, 510–548 (2017).
49. Rapoport, A. Mathematical models of social interaction. in *Handbook of Mathematical Psychology* Vol. 2 (eds. Galanter, R. A., Lace, R. R., & Bush, E.) 493–580. (John Wiley & Sons, 1963).
50. Rambaran, J. A., Dijkstra, J. K., Munniksma, A. & Cillessen, A. H. The development of adolescents' friendships and antipathies: a longitudinal multivariate network test of balance theory. *Soc. Netw.* **43**, 162–176 (2015).
51. Harary, F. A structural analysis of the situation in the Middle East in 1956. *J. Confl. Resolut.* **5**, 167–178 (1961).
52. Neal, Z. P. A sign of the times? Weak and strong polarization in the US Congress, 1973–2016. *Soc. Netw.* **60**, 103–112 (2020).
53. Doreian, P. & Mrvar, A. Structural balance and signed international relations. *J. Soc. Struct.* **16**, 1–49 (2015).
54. Szell, M. & Thurner, S. Measuring social dynamics in a massive multiplayer online game. *Soc. Netw.* **32**, 313–329 (2010).
55. Belaza, A. M. et al. Social stability and extended social balance—Quantifying the role of inactive links in social networks. *Phys. A: Stat. Mech. Appl.* **518**, 270–284 (2019).
56. Facchetti, G., Iacono, G. & Altafini, C. Computing global structural balance in large-scale signed social networks. *Proc. Natl Acad. Sci.* **108**, 20953–20958 (2011).
57. Liu, H., Qu, C., Niu, Y. & Wang, G. The evolution of structural balance in time-varying signed networks. *Future Gener. Comput. Syst.* **102**, 403–408 (2020).
58. Askarisichani, O. et al. Structural balance emerges and explains performance in risky decision-making. *Nat. Commun.* **10**, 2648 (2019).
59. Hintzman, D. L. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol. Rev.* **95**, 528 (1988).
60. Gigerenzer, G. How to explain behavior? *Top. Cogn. Sci.* **12**, 1363–1381 (2020).
61. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
62. Flache, A. et al. Models of social influence: towards the next frontiers. *J. Artif. Soc. Soc. Simul.* **20** (2017).

Acknowledgements

This study was supported by the Complexity Science Hub Vienna and the Santa Fe Institute.

Author contributions

All authors have conceived and conducted the study, analyzed the data, and written the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44260-024-00022-y.

Correspondence and requests for materials should be addressed to Mirta Galesic or Henrik Olsson.