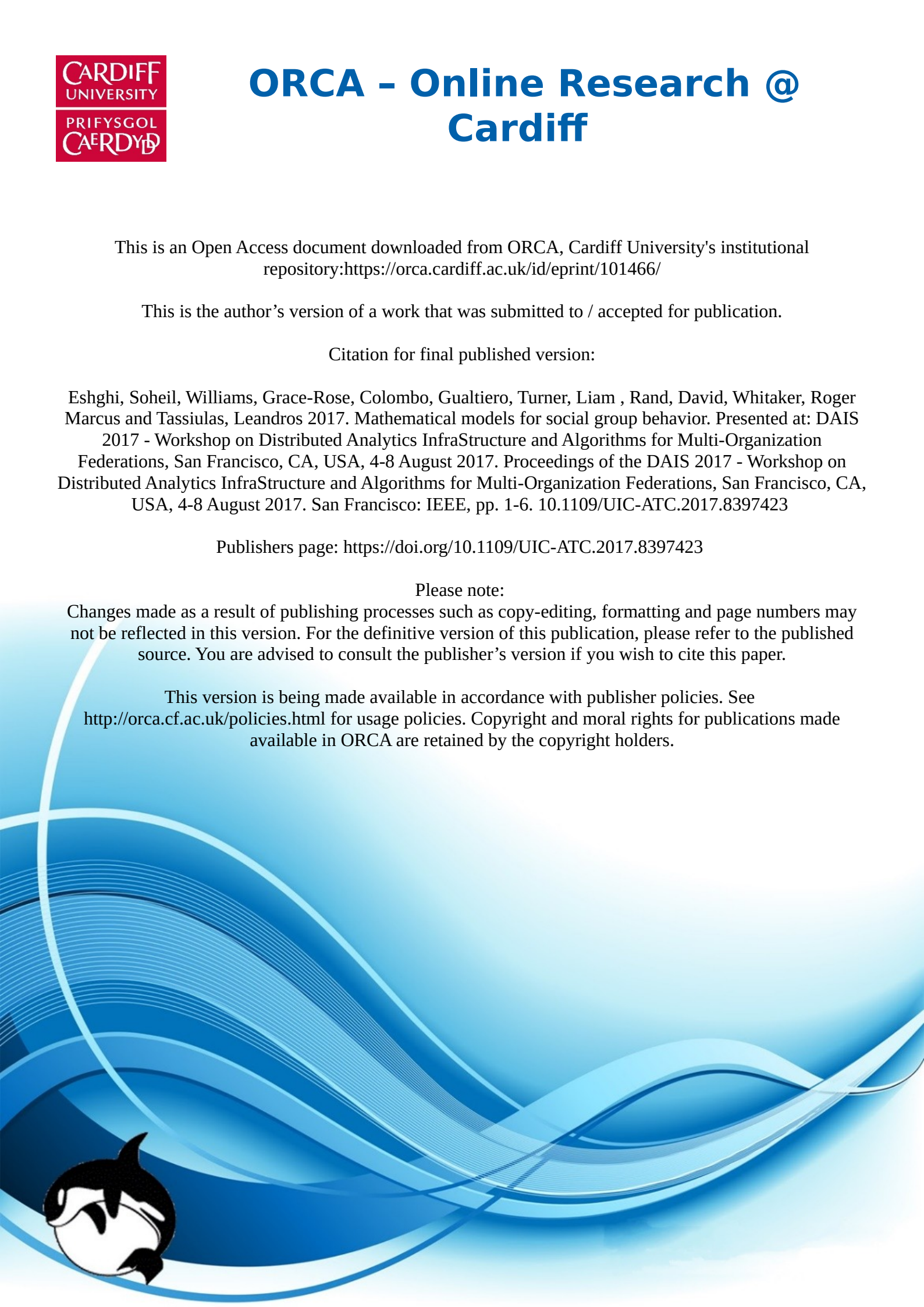

Mathematical models for social group behavior

Soheil Eshghi*, Grace-Rose Williams†, Gualtiero B. Colombo‡, Liam D. Turner‡,
David G. Rand§, Roger M. Whitaker‡, Leandros Tassiulas*
* *Yale Institute for Network Science (YINS) and Electrical Engineering Department,*
Yale University, New Haven, USA
E-mail: {soheil.eshghi, leandros.tassiulas}@yale.edu
† *Defence Science and Technology Laboratory (Dstl),*
Porton Down, UK
E-mail: grwilliams1@dstl.gov.uk
‡ *School of Computer Science & Informatics,*
Cardiff University, Cardiff, UK
E-mail: {ColomboG, TurnerL9, whitakerrm}@cardiff.ac.uk
§ *Psychology Department & Economics Department & Yale School of Management (YSOM),*
Yale University, New Haven, USA
E-mail: david.rand@yale.edu

***Abstract*—In this paper, we seek to identity how mathematical and economic analysis can be used to gain insights about the mutation of social groups. Group mutability has been studied in multiple domains, with insights generated on significant factors at differing scales. Mathematical modeling enables the simultaneous study of such phenomena, understanding interactions and generating hypotheses for experiments. In particular, we focus on group fracture, where individuals leave groups of which they are members. For example, this can be due to perceived differences with other group members due to norm related conflict (such as extreme actions by some members). Our aim is to consider simple mathematical models incorporating a selection of social and psychological theory which describes these phenomena as a way to understand their interplay, and describe the trade-offs and challenges. This will help a federation model the behavior of extremist groups, and determine not only when an intervention is necessary, but also the best course of action to take to induce the fracture of such groups. This paper is an exploratory investigation into methods of achieving this goal and evaluating the usefulness of the outputs to federations.**

***Keywords*-group dynamics; social groups; modeling;**

I. Introduction

The stability of groups is an emerging topic of study in differing contexts (see [1] for a summary). While some interest is focused on the evolution of cooperation [2], [3] and groups in a biological and evolutionary setting, there is significant interest in understanding the predictors of group stability and mutation in management and counterterrorism [4] settings. The effect of different phenomena on group behavior and dynamics, across individual to group scales, has been the subject of much mathematical study since the 1940's [5] which has typically employed methods from physics and game theory. The physics-inspired approaches provide insight mainly through simulation (e.g., [6], [7]), and involve many parameters, while the game theoretic models either directly focus on simple two-player ultimatum and prisoner's dilemma games [8], [9], [10], [11], [12], or bespoke games relevant to particular scenarios (e.g., group conflict [13], [14], [15]). In the intervening time, many micro-economic models have been developed to describe specific group-dependent actions [16], [17], [18], [19]. Recently, Kranton [20] provided a roadmap for the integration of these models to create a meta-model for group behavior. In this paper, we use this roadmap to develop a mathematical theory of group fracture and stability. This will help us to identify stable groups and their most vulnerable/reluctant group members, and serve as a basis for reasoning about the interactions of the aforementioned phenomena. We categorize these phenomena based on their social dependence: interdependent phenomena are dependent on the choices and actions of other group members, while independent phenomena are related to the psychology and perception of the individual irrespective of other group members.

A. Interdependent group-based phenomena

It has been observed that individuals exhibit cooperative, sometimes costly behavior to maintain groups [2] in differing contexts, even if the gains (material, reputational, or otherwise) are not immediate [21], [22], [23]. These behaviors may be due to personal normative beliefs or morals, or descriptive or injunctive norms. The defining line between these two types of behavior is the conditional preference for the action: personal beliefs and morals may persist without any type of reciprocity or expectation of others (i.e, they are independent), while the effect of descriptive or injunctive norms would be lessened without social feedback on their enforcement (i.e., they are interdependent). Knowledge of group norms, social rules which define what behavior is expected of an individual within the group as a condition of membership [23], [24], and their effects is critical in understanding the stability and fracture of groups.

Technically, for a behavior to be considered a social norm, there need to be both empirical and normative expectations (expectations about how others behave, and how the individual believes they are expected to behave by others), as well as conditional preference (the behaviors critically depend on their being mirrored by a social reference group) [25]. These norms impose a burden on some group members, which may itself be a source of tension for the individual (if they are onerous) or the group itself (if they are unfair).

However, norms are not the only interdependent group-based phenomena that may affect group mutability. The choices of other members of the reference group may have indirect effects on the choices of a group member (i.e., act as externalities) through processes such as social comparison.

Social comparison is a mechanism through which individuals compare their opinions and actions with others to gain a better and possibly accurate self-evaluation [26]. Festinger [26] originally hypothesized that individuals compare themselves against similar people. One such case would be comparison against in-group members (e.g., comparing wages against your peers versus some other occupation). The exact effect of a particular comparison can depend on context [27] or group-membership [28], and it can be unidirectional (i.e., only considering those better-off or worse-off in the comparison) [29], [30] or multi-directional [31]. This internal mechanism both affects individual decision-making and, indirectly, the decision-making of others [16], [32], [33].

Social comparison among intrinsically similar individuals may lead to the straightforward adoption of successful behavior [34]. However, with heterogeneity in abilities, observing the actions of others is not necessarily informative of their effort (i.e., their strategy). Thus, the effect of social comparison is a comparison of observable actions/behaviors, or rewards, with other group members.

Under a multi-directional model of social comparison, any comparison between two individuals makes one better off and the other worse off. Those that will fare favorably in the comparison have a natural incentive to compare themselves against the in-group. Thus, the question arises as to how one can explain why individuals who might receive a negative result from this comparison, continue to remain in groups (without assuming a preference for group membership come what may). This work seeks to answer this question by considering the inter-play of this phenomenon with other that manifest at different time and population scales.

B. Independent group-based phenomena

Belonging to a group cannot just be explained through a transactional view of group norms. Individuals may be drawn to groups for many other reasons, which have been extensively studied in the social and cognitive psychology literature. While examining norms focuses on the interdependent underpinnings of the group, many theories have been postulated to explain non-reciprocal ("independent") reasons for identifying with a group. Self-categorization theory is one such influential theory.

According to self-categorization theory [35], individuals can characterize themselves in different ways, with the salience of a categorization to the individual forming the basis of de-personalization, whereby the individual sees themselves as an "interchangeable exemplar of a social category" [36], and thus accept the norms of the social group. The salience of a group to an individual depends on the unconscious process of accentuation, whereby differences with other groups and similarities within the group are amplified. Given the multiplicity of reasons affecting this accentuation, it is not unreasonable to assume that there will be a significant amount of heterogeneity in a social group as to extent to which each member categorizes themselves as part of the group, and therefore is reluctant to leave it.

While self-categorization codifies "attachment" to the group, other factors may cause the same effect of keeping members within a group (i.e., acting as friction or stickiness which prevents individuals leaving the group). There may exist perceived, implicit, or explicit threats to individuals who leave a group (e.g., violence, ostracism), either from other group members or from the out-group (making this an independent or interdependent process depending on the case). This may cause reluctance on the part of the individual, keeping them in the group even in the face of onerous norms.

C. Research Question

Any model of group mutability should incorporate multiple social and psychological theories, including ones that describe the relation of the self to the group (self-categorization, social identity, etc.) and those that describe the relation of the group to its constituent individuals (group norms, group cohesiveness, etc.). However, given the complexity of human interactions, the wealth of theory, and the differences in context, there is a trade-off between the mathematical tractability of a model and how closely it captures human interactions. Furthermore, and more broadly, the question of how to integrate information about individual behavior in groups into a model of group-level behaviors has surprisingly been relatively neglected by previous literature.

In particular, we focus on the effect of norm-related conflict in the fracture of a group under these conditions. This conflict can manifest when empirical expectations of individuals are in conflict with normative expectations, leading to the supremacy of empirical expectations and the changing of normative expectations [25]. It can also manifest when a specific social norm puts a significant strain on a particular individual (conflicting with a fairness norm), or when the burden it places on an individual is significant enough to convince them

to leave the group (and to risk the associated negative consequences).

D. Contribution

In this paper, we propose new analytical approaches that aim to incorporate some of the above features into a mathematical model, by focusing on the tensions between an individual and their relationship with the group to which they belong and how this influences the stability or fracture of the group. We seek to elucidate these tensions and trade-offs mathematically, combining these factors so that we can capture conditions that lead to changes in the condition of a group. Specifically, we are interested in the stability or otherwise of groups of non-state actors [4].

As an example, this paper will show the linkage between mathematical theory and social phenomena through which group stability can be assessed by way of representations of utility. Characterizing the stability of groups mathematically has the side-benefit of allowing a quantification of a group-member's relative attachment to the group. This is especially important to characterize for internally-stable extremist groups, where the coalition may seek to tactically target particular individuals with incentives to leave the group (e.g., monetary incentives, information campaigns) so as to create division within their structure.

It should be noted that while we borrow liberally from the underlying assumptions of many of the underlying models (as will be stated), our modeling approach is distinct, especially, from the game-theory literature which considers repeated interactions modeled as simple games, as our focus is on the inter-play of many different phenomena at different scales. This is important due to the fractal nature of modern knowledge, especially so as group mutability has been studied in such varied communities and a meta-model will allow the simultaneous exploitation of diverse insights, as well as facilitating the understanding of their interplay.

II. Mathematical model

In this work, we develop utility models for individuals that incorporate the group-based phenomena described above. These utilities will have three broad parts:

1) *Intrinsic*: This is a simple cost-benefit calculation that determines the effort the individual expends in group-related activities. The calculation of actions is complicated by the diverse abilities of individuals.

2) *Externalities*: The actions chosen by other group members affect the perceptions of the individual and their choice of effort/action. Thus, the actions chosen by others act as an externality that modifies the utility, and possibly the chosen action of an individual. This is complicated by the limited observations of other group members and the difficulty in inferring the reason behind their actions.

3) *Group-based effects*: While the previous two parts of the utility are related to individual interactions, there are group-based effects that manifest on longer timescales, e.g. a utility due to attachment to groups (related to the salience of the group), and a representative utility related to norm-based interactions with other group members, which may decrease the utility of a single individual to the the benefit of other group members. More precisely, this utility models normative expectations by the individual.

We map these types of utility into the three broad categories outlined by Kranton [20], which are dubbed the *short*, *medium*, and *long run*. In her framework, individuals choose actions in the *short run* taking expectations, norms, identities, and categories to be fixed. In the *medium run*, individuals can take some actions to modify their empirical and normative expectations (to resolve conflict) or their relative attachment to groups (categorization). In the *long run*, nothing is fixed.

In the short run, we identify social comparison [26] as one of the externality-causing phenomena and use the mathematical framework set up by Clark and Oswald [16]. We will also incorporate a heterogeneity of individuals in terms of abilities. One of the interesting results of such a framework is that it has been shown to implicitly model both conventional and contrarian characteristics in individuals [16].

In the medium run, we will consider empirical and normative expectations [25] and group norms [37] and their effect on group members. Normative expectations act as a belief about expected behavior of the individual, while empirical expectations act as an expectation of future behavior by others. When they are in conflict, the conditional preference property of group norms may make an individual less likely to follow them, adopting the empirical norms instead. The various sources of group friction/stickiness (e.g., self-categorization) are viewed in aggregate as a measure of how willing an individual is to suffer onerous norm-related actions/punishments. The questions investigated in this time horizon are whether a norm is self-consistent (e.g., will empirical expectations match normative expectations) and whether it is unfair [38]. If the answer to any of these questions is no, the one can predict that the normative expectations will evolve in some way in the long-run to resolve these conflicts (in other words, the description of the norm is not stable). Another question of interest is whether the burden placed on an individual is related to their identification with the group. This approach is inspired by Bicchieri [25].

The primary question under investigation in the *long run* is whether a given group is stable under the evolved (and thus self-consistent) normative expectations of the medium-run. For the purpose of this study, we define stability to mean that no member of the group would be incentivised to leave the group. We use the model of

rational agents with clear preferences used in economics in this definition. If it is indeed stable, and no member will disassociate with the group of their own volition, we are interested in understanding which member is the most vulnerable group member to target with an incentive to facilitate their leaving of the group.

After making the case for a model for group stability that considers social comparison and group norms in the next subsections we describe the constituent parts of the mathematical model and its underpinnings from an analytic perspective.

A. Short Run

We now present the additive social comparison model courtesy of [16] for the short run. In this model, individuals choose how much effort to put into an action (that is related to the purpose of the group) given their ability in that task, which is a personal, unobservable, and heterogeneous trait, as well as subjective social comparison. In this model, an individual's private ability/fitness to perform tasks related to that goal is captured by their type $\theta \in [\underline{\theta}, \overline{\theta}]$. We assume this parameter does not change in the time-scale of consideration. An individual considers their type in choosing their effort which leads to their (observable) action $a \geq 0$.

Individuals are rarely rewarded for unobservable effort - rather, outcomes typically depend on actions. Thus, there is a trade-off for each individual, between the rewards related to taking an action, and the cost of the effort it requires. This internal trade-off is further complicated by the psychological effects of social comparison. This is the central question of the short-run model.

The utility that an individual derives from comparison against their reference group depends on a characterization of the group's actions, e.g. via the mean observed action a^*. To estimate this value accurately, individuals must accumulate information, therefore this representative action may not always align with the true population mean. However, as an individual encounter more and more people, the empirical mean observed action will converge to the mean of the distribution. In this time-frame, group norms and identities can be assumed to be fixed.

B. Medium run

In the medium run, we focus on group effects. We assume that the short run dynamics have reached an equilibrium, such that the perception of a^* by group members has converged to the real population average, and individuals receive a utility of $u_{sr}(\theta)$ from the short-run dynamics.

The group has a salience to an individual that is captured through a parameter $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. This parameter can also model how important a group is to a person's self-concept, or, inversely, how difficult a group is to leave (i.e., what adverse consequences would result from such an action). In effect, this acts as a "friction" term that keeps group members inside the group. For example, for a minimal group, γ would be small, while it would be large for a group that is especially important to an individual's self-concept (as described in identity fusion theory). We assume that this parameter is fixed in the short and medium run. Thus, we can assume that there is a probability density function $P(\theta, \gamma)$ over the set of (θ, γ) pairs which describes the population. Note that this allows there to be a possible correlation between ability and salience of the group to the individual.

Each individual, knowing their private ability and the salience of the group to them, takes the action they believe other group members expect someone in their situation to take (normative expectations). We consider quantify the preferences of the individuals over these norm-related actions through a function, $b_n(\theta, \gamma)$, that considers the net effect of these actions on the individual. For example, some norms may involve some group members helping other in-group members. In these circumstances, adhering to normative expectations may not just place no burden on the individual being helped, they might also decrease the effort they need to exert in completing the task. On the other hand, the additional burden on the helpers/donors may make group membership less desirable to them.

Each individual will also have expectations about norm-related behavior from other group-members. These expectations will be empirical [25], and will align with their observed behavior. We quantify the preference of an individual over these empirical actions through the function $b_e(\theta, \gamma)$, which signifies the understanding of the individual about the actions taken by other group members given their private information (translated to the same scale as $b_n(\theta, \gamma)$).

There may arise a case where an individual's empirical expectations conflict with their normative ones. This will be the case when the behavior they observe from others is incompatible with the behavior they deem the others expect from them. Under these conditions, the normative expectation will be amended over time to be compatible with the empirical expectation [25]. Since we are not explicitly concerned with the process under which these changes happen, we consider the end-result, whereby we have a convergence of b_n and b_e to a compatible functional description of the norm, $b(\theta, \gamma)$. This would be the shared norm that is enforced by the group in the medium run, possibly through sanctions [39]. While there is significant work in understanding sanctioning decisions and their effect on norms, we do not consider them explicitly in this framework.[1]

For this norm to be self-compatible, it must be possible for each individual in the group to perform the action which they believe the norm prescribes for them, and

[1] In the long-run, we discuss the fact that leaving the group would result in the loss of the benefit gained from group membership (codified in our model through γ). This loss can capture the effect of ostracization and sanctions for not abiding by norms.

for other group members to be able to sustain the expectations of an individual.

C. Long run

In the medium run, we assumed that individuals have "learned" the norms of the group (e.g., perceived norms of group members and collective norms have converged). In the long-run, given that even the group itself is not considered fixed, we focus on the stability of the group. In particular, we discuss each individual's choice of whether to remain in the group and to be bound by its norms, or to leave the group and risk sanctions by group members. The preference is codified through a comparison of the individual's utility within the group and outside the group.

Leaving the group would also modify the expected short-run and medium-run utility of the individual, so as in the short-run, and outside the group, the individual will be deprived of the feedback provided by observing the actions of group-members. This would translate to a change in an individual's utility function, and therefore their chosen action. The medium-run effects we described are both related to group membership, and will have no effect once the individual leaves the group[2].

1) Long run group stability: Individuals make the decision to leave the group or to stay by considering their preference over these choices, as captured by a comparison of total in-group utility with that they could expect to sustain outside the group in the short and medium run:

$$u_{mr}(\theta,\gamma) + u_{sr}(\theta) \geq u^o_{mr}(\theta,\gamma) + u^o_{sr}(\theta), \tag{1}$$

where $u_{sr}(\theta)$ and $u_{mr}(\theta,\gamma)$ (respectively $u^o_{sr}(\theta)$ and $u^o_{mr}(\theta,\gamma)$) are an individual's expected short-run and medium-run utility inside (outside) the group. Note that the difference between the two sides of the inequality represents the starkness of the difference between the choices for (θ,γ)-individuals. One can think of this difference as representing the additional encouragement (in the form of outside incentives) the individual would need to be convinced to leave the group (such as the rewards offered to members of terrorist groups to facilitate their de-radicalization [40]).

We can now provide a mathematical definition for long-run *stability* $\mathscr{S}$ of a group: $\mathscr{S}$ is the minimum additional incentive that has to be offered to group members to cause one of them to leave in the long run:

$$\begin{aligned} \mathscr{S} := \min_{k\geq 0} \quad & k \\ \text{s.t.} \quad & u_{mr}(\theta,\gamma) + u_{sr}(\theta) - u^o_{mr}(\theta,\gamma) - u^o_{sr}(\theta) \leq k \quad \exists_{\theta,\gamma}. \end{aligned} \tag{2}$$

Notice that if (1) does not hold for an individual in the group, then by default $\mathscr{S} = 0$. In such cases, one can use this framework to study how many individuals would have to leave the group to make it stable, potentially via computational simulations.

In the long run, we seek to study what type of shared norms maximize the stability of a groups. In this study we are not investigating how these norms are generated nor the exact mechanisms by which they are maintained, but only on their effect on the mutability of the group.

One could also add other constraints on the group norm that align with theory. For example, it has been argued that a complementary measure of fairness of a norm is required to capture effects such as inequality aversion. One such constraint could penalize norms that place large expectations, or give large benefits, to a subset of people.

III. Conclusions, Challenges, and Future Work

In this paper, we have described a mathematical framework for social group mutability. We show how various interdependent and independent processes that have been seen to affect group mutability can be captured in an economics-inspired mathematical model. This is a significant step in understanding the possible interactions between these phenomena that may be hard to see in laboratory experiments. Understanding this interplay mathematically will allow us to posit experimentable hypotheses about these interactions.

In future work, we will use this framework to integrate specific mathematical models for group mutability-relevant phenomena, and create optimization-based models of individual behavior. While the outcome of these models typically depends upon socially-inspired parameters, it is often difficult to estimate these parameters from data. This is both due to the crudeness of the models and the unavailability of enough data for model fitting. However, such estimation and fitting is necessary to generate exact hypotheses for specific experiments, and is thus an important step in our future work.

Acknowledgment

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

References

[1] W. J. Wildman and R. Sosis, "Stability of groups with costly beliefs and practices," *Journal of Artificial Societies and Social Simulation*, vol. 14, no. 3, p. 6, 2011.

[2] H. Gintis, E. A. Smith, and S. Bowles, "Costly signaling and cooperation," *Journal of theoretical biology*, vol. 213, no. 1, pp. 103–119, 2001.

[2]This is due to the way we have encoded γ. One can equally plausibly define γ to emphasize the relative dis-utility of being in the out-group.

[3] A. Bear and D. G. Rand, "Intuition, deliberation, and the evolution of cooperation," *Proceedings of the National Academy of Sciences*, vol. 113, no. 4, pp. 936–941, 2016.

[4] S. G. Jones and M. C. Libicki, *How terrorist groups end: Lessons for countering al Qa'ida*. Rand Corporation, 2008.

[5] K. Lewin, "Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change," *Human relations*, vol. 1, no. 1, pp. 5–41, 1947.

[6] K. Carley, "A theory of group stability," *American sociological review*, pp. 331–354, 1991.

[7] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, 2007.

[8] S. M. Allen, G. Colombo, and R. M. Whitaker, "Cooperation through self-similar social networks," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 5, no. 1, p. 4, 2010.

[9] D. Nettle and R. I. Dunbar, "Social markers and the evolution of reciprocal exchange," *Current Anthropology*, vol. 38, no. 1, pp. 93–99, 1997.

[10] R. M. Axelrod, *The evolution of cooperation*. Basic books, 2006.

[11] D. Hales and B. Edmonds, "Applying a socially inspired technique (tags) to improve cooperation in p2p networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, no. 3, pp. 385–395, 2005.

[12] M. A. Nowak, "Five rules for the evolution of cooperation," *science*, vol. 314, no. 5805, pp. 1560–1563, 2006.

[13] D. Verma, G. Pearson, D. Felmee, A. Verma, and R. M. Whitaker, "A generative model for predicting terrorist incidents," in *SPIE Defense + Security Symposium: Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII*, 2017.

[14] R. M. Whitaker, L. D. Turner, G. Colombo, D. Verma, D. Felmee, and G. Pearson, "Intra-group tension under inter-group conflict: a generative model using group social norms and identity," in *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics*, 2017.

[15] A. B. Naugle and M. L. Bernard, "Using computational modeling to examine shifts towards extremist behaviors in european diaspora communities," in *Advances in Cross-Cultural Decision Making*. Springer, 2017, pp. 321–332.

[16] A. E. Clark and A. J. Oswald, "Comparison-concave utility and following behaviour in social and economic settings," *Journal of Public Economics*, vol. 70, no. 1, pp. 133–155, 1998.

[17] C. F. Manski, "Economic analysis of social interactions," National bureau of economic research, Tech. Rep., 2000.

[18] G. A. Akerlof and R. E. Kranton, "Economics and identity," *The Quarterly Journal of Economics*, vol. 115, no. 3, pp. 715–753, 2000.

[19] J. Elster, "Social norms and economic theory," in *Culture and Politics*. Springer, 2000, pp. 363–380.

[20] R. E. Kranton, "Identity economics 2016: Where do social distinctions and norms come from?" *The American Economic Review*, vol. 106, no. 5, pp. 405–409, 2016.

[21] T. Yamagishi and K. S. Cook, "Generalized exchange and social dilemmas," *Social Psychology Quarterly*, pp. 235–248, 1993.

[22] T. Watanabe, M. Takezawa, Y. Nakawake, A. Kunimatsu, H. Yamasue, M. Nakamura, Y. Miyashita, and N. Masuda, "Two distinct neural mechanisms underlying indirect reciprocity," *Proceedings of the National Academy of Sciences*, vol. 111, no. 11, pp. 3990–3995, 2014.

[23] E. Fehr and S. Gächter, "Altruistic punishment in humans," *Nature*, vol. 415, no. 6868, pp. 137–140, 2002.

[24] R. Axelrod, "An evolutionary approach to norms," *American political science review*, vol. 80, no. 04, pp. 1095–1111, 1986.

[25] C. Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press, 2016.

[26] L. Festinger, "A theory of social comparison processes," *Human relations*, vol. 7, no. 2, pp. 117–140, 1954.

[27] P. Lockwood, C. H. Jordan, and Z. Kunda, "Motivation by positive or negative role models: regulatory focus determines who will best inspire us." *Journal of personality and social psychology*, vol. 83, no. 4, p. 854, 2002.

[28] R. L. Collins, "For better or worse: The impact of upward social comparison on self-evaluations." *Psychological bulletin*, vol. 119, no. 1, p. 51, 1996.

[29] T. A. Wills, "Downward comparison principles in social psychology." *Psychological bulletin*, vol. 90, no. 2, p. 245, 1981.

[30] B. P. Buunk, R. L. Collins, S. E. Taylor, N. W. VanYperen, and G. A. Dakof, "The affective consequences of social comparison: either direction has its ups and downs." *Journal of personality and social psychology*, vol. 59, no. 6, p. 1238, 1990.

[31] M. B. Brewer and J. G. Weber, "Self-evaluation effects of interpersonal versus intergroup social comparison." *Journal of personality and social psychology*, vol. 66, no. 2, p. 268, 1994.

[32] G. Roels and X. Su, "Optimal design of social comparison effects: Setting reference groups and reference points," *Management Science*, vol. 60, no. 3, pp. 606–627, 2013.

[33] R. M. Whitaker, G. B. Colombo, S. M. Allen, and R. I. Dunbar, "A dominant social comparison heuristic unites alternative mechanisms for the evolution of indirect reciprocity," *Scientific Reports*, vol. 6, 2016.

[34] D. Friedman, "On economic applications of evolutionary game theory," *Journal of Evolutionary Economics*, vol. 8, no. 1, pp. 15–43, 1998.

[35] J. C. Turner and P. J. Oakes, "Self-categorization theory and social influence." 1989.

[36] J. C. Turner, "Social categorization and the self-concept: A social cognitive theory of group behavior," *Advances in group processes*, vol. 2, pp. 77–122, 1985.

[37] D. C. Feldman, "The development and enforcement of group norms," *Academy of management review*, vol. 9, no. 1, pp. 47–53, 1984.

[38] I. Castelli, D. Massaro, C. Bicchieri, A. Chavez, and A. Marchetti, "Fairness norms and theory of mind in an ultimatum game: judgments, offers, and decisions in school-aged children," *PloS one*, vol. 9, no. 8, p. e105024, 2014.

[39] E. Fehr and U. Fischbacher, "Social norms and human cooperation," *Trends in cognitive sciences*, vol. 8, no. 4, pp. 185–190, 2004.

[40] M. Crenshaw, "Theories of terrorism: Instrumental and organizational approaches," *The Journal of strategic studies*, vol. 10, no. 4, pp. 13–31, 1987.