

# Covert Subversive Agents and Consensus Disruption on Large Projects

Johannes Vorster and Louise Leenen

University of the Western Cape, South Africa & Centre for Artificial Intelligence Research (CAIR), South Africa

[jvorster@uwc.ac.za](mailto:jvorster@uwc.ac.za)

[lleenen@uwc.ac.za](mailto:lleenen@uwc.ac.za)

**Abstract:** Is it possible to dramatically affect and influence military and other projects through social engineering of the consensus processes? In this paper we explore the impact that subversive agents can have on the ability of projects to move forward by disrupting the social cohesion and decision-making abilities of the processes designed to reach consensus. A consensus simulator is used to model group social cohesion behaviour in the context of project deliverables and show what the effect can be on the effort to reach consensus (number of meetings) as well as the time to reach consensus (calendar time) when subversive agents attempt to influence the groups making up the project team in such a way that it delays the ability of the team to reach consensus on key decisions. Many military options are available to delay enemy projects, including the assassination of enemy scientists, sanctions aimed at denying key project components, or even direct military action such as bombing the enemy facilities. However, this paper focusses on aspects of soft-force projection through covert disruption of project timelines. A social simulator was constructed that models individual agent's beliefs about various key topics within the context of a project. The effect that a small number of subversive agents can have on the time- and effort of a project is shown. In their covert actions, these subversive agents need to stay hidden, and thus their covert actions are limited, yet they can exert significant damage to the project in terms of delays. In this paper we present results showing the effects that such a small group can have, as well as pointing out that there seem to be a critical group size over which the subversive agents can not only have significant impact on project-delays but can also steer and direct certain key decisions.

**Keywords:** Subversive agents, Consensus, Consensus simulation, Multi-agent simulation, Stochastic consensus models

---

## 1. Introduction

Achieving consensus among diverse stakeholders is essential for mission success in large projects as well as military operations (Floyd, 1992; Charchian, 2001). However, this can be challenging in complex and dynamic environments where subversive agents (SAs) may intentionally disrupt the consensus-building process. For the purposes of this paper SAs are individuals or groups who promote views that cause more debate and delay the achievement of consensus and they dynamically change their views to purposefully delay the consensus process. It is important to distinguish between this subversive behaviour and normal healthy debate and disagreements, here the subversive agents promote views that they themselves may not believe, but they do so to create discord.

This paper explores the impact of SAs on consensus-seeking processes using a multi-agent simulator (MAS) and aims to provide insights into how organizations can mitigate their effects. The findings of this study have implications for Information Operations (IO) in military contexts, where achieving consensus among different stakeholders is critical for effective decision-making. The exploitation of project related social cohesion by subversive agents to affect project delivery has not been studied, to your knowledge.

### 1.1 Historic Perspective

IO involves the integration of various information-related capabilities to influence, disrupt, corrupt or usurp an adversary's decision-making process while protecting one's own. The presence of SAs can significantly impact IO efforts by delaying or disrupting consensus-building processes, leading to ineffective decision-making and potentially compromising mission success.

Therefore, understanding how SAs operate and their impact on consensus-building processes is critical for military planners and decision-makers involved in IO efforts. This paper builds on our previous research on consensus-seeking processes using MAS (Vorster & Leenen, 2023a, 2023b) and provides insights into how organizations can identify and mitigate the effects of SAs in IO contexts.

There have been several instances in military history where projects were delayed due to foreign influence, leading to inter-team conflict and discord. One such example is the development of the US F-35 Joint Strike Fighter program, which was delayed due to foreign influence from partner countries such as the United

Kingdom and Italy (Maldonado, 2015). These countries had specific requirements for the aircraft that were not initially included in the program, leading to disagreements and delays in the development process.

Another example is the development of the US M1 Abrams tank, which was delayed due to foreign influence from Germany, which had specific requirements for the tank's armour and weaponry (McDonald, 1995). This led to disagreements and delays in the development process, causing tension between the US and German teams working on the project. Other examples include the development of the Joint Strike Fighter (JSF)

These examples were not discord created by adversaries, but we can imagine that such conflict can be created by agents influenced by adversaries.

### **1.2 Healthy Conflict Within Project Teams**

Earlier case studies of project failures have shown repeatedly that the top causes are project complexity and the failure of project team members to reach consensus on key points, be it requirements, design, or implementation aspects (Al-Ahmad et al., 2009; Whitney & Daniels, 2013; Kian, Sun, & Bosché, 2016; Waheeb & Andersen, 2022).

The process of reaching consensus implies that functionally diverse teams must understand the problem, surface, and share knowledge about the problem, identify any commonalities that may exist between views as well as differences in views, consolidate every difference in perspective, and through a team effort propose a solution that incorporates the final views of the whole team (Cheung et al, 2016). However, studies have also shown (eg Enyinda et al 2022) that healthy task conflict can increase a team's ability to reach consensus. This positive correlation between task conflict and team performance is especially notable in sufficiently complex tasks such as when there is no clear solution and involve significant uncertainty, such as during innovation and design, so that subjective factors determine the consensus outcome (Paletz et al 2017), neither do all problems have a single solution.

### **1.3 Subversive Agents Versus Healthy Conflict**

Some remarks are needed on subversive agents versus conventional disagreement on topics between peers trying to find satisfactory solutions. In many projects the issues are complex, and people have justifiable reasons for arguing positions that others may consider unimportant. This is normal, and part of finding good solutions. The fact that there are debates imply that topics are not simple. Good communication, collaboration, and social skills within the milieu of task conflict has been found to be important for a team to reap the benefits from healthy conflict (Hirvonen, 2019). The rhetorical question "is there a way to tell the difference?", is an open question.

Debates within this context leads to longer times to reach consensus and to reject healthy conflict would lead to worse decisions being made. That is, the conflict creates points of debate that ultimately leads to better decisions. In this paper we use the term *subversive agent* to mean someone that is proactively trying to find arguments, trying to convince others of notions that they themselves do not believe in, specifically with the purpose of causing project delays and derailing decisions.

### **1.4 Social Influence Models**

Early models of social influence were based on the effects that individuals with committed views were on the collective. A *committed view* refers to a view that is unchanging in the face of alternative views. Individuals with committed views have an unwavering belief in their opinion and they retain their view despite other individuals having other views. Over time, these individuals will convince others of their view, but since they are committed to their view, they will not change their position. So, the expected outcome is that eventually this committed view will take over as the dominant view.

Consider the following two example, the theory that the earth goes around the sun, as opposed to the over way around; and the spread of the conviction that female contraceptives should be used. There are many other examples the reader could identify, from the use of microwaves to the belief that smallpox vaccinations are good.

In both these cases the majority is convinced of something and then a new idea starts to propagate, driven by individuals with unwavering conviction.

We briefly mention two early influential models namely the *Bass (1969) diffusion model* and later *Granovetter's (1978) threshold model*. The diffusion model is based on product adoption, e.g. hairspray, where

the probability of an individual taking up the product is linear to previous buyers of the product. Adoption is driven by the number of other people that can be seen as having adopted. In the threshold model individuals observe the behaviour and views of their social neighbours, and if a certain threshold percentage is seen to have switched to the new view, then they will switch as well. The consequences of these models and their relevance to real-world problems has been studied extensively due to the importance of this topic to political and social views.

It was shown that there exists a critical threshold at about 10% of the population. If the number of individuals within the population that believe the new view exceeds 10%, then the time for the rest to be convinced is traumatically decreased (Xie et al, 2011). Later, Iacopini et al (2022) studied consensus processes through groups. In their model groups will gather and listen to a single speaker. They define a probability (beta,  $\beta$ ) that the listeners will be convinced, as a group, and if that fails, everyone still has their own probability (alpha,  $\alpha$ ) that they may or may not be convinced. Let us call  $\beta$  the group's gullibility (our term, not theirs). They found that the ability of the committed minority to convince the entire population of their views falls into three distinct behavioural regions based on how gullible of the groups. Let us say that the initial majority view is represented by *view-a*, and the committed minority holds a view represented by *view-b*.

In the first region, with low  $\beta$ , the final equilibrium number of people with *view-b* is not the full population. If  $\beta$  is very small, then the equilibrium number of individuals that hold *view-b* is small. As  $\beta$  increases, the equilibrium number of individuals with *view-b* increases dramatically. At a specific critical  $\beta$  the equilibrium position reached is for the entire population to hold *view-b*. This ushers in region two where the minority can convince the entire population. As  $\beta$  is further increased, a second critical value is reached. For  $\beta$  values beyond this point the majority retain their *view-a*, and the minority is isolated to be the only individuals holding *view-b*.

### **1.5 Multi-Agent Systems in the Study of Consensus**

We now turn to the more technical discussion on the use of Multi-Agent Systems (MAS). The past decade has seen significant studies on MAS and how to achieve and maintain consensus with a dynamically changing group of such agents under various conditions. For a detailed overview of this complex topic see reviews by Bao et al (2022) and Yang et al (2022). The applications of this research have been in high-speed consensus for Unmanned Aerial Vehicles (UAVs) and autonomous ground vehicles. One of the significant results is that if the difference of views between the various agents are measured, then that number decreases exponentially over time as the consensus protocol used drives agents to adopt the same views (Wei, 2021).

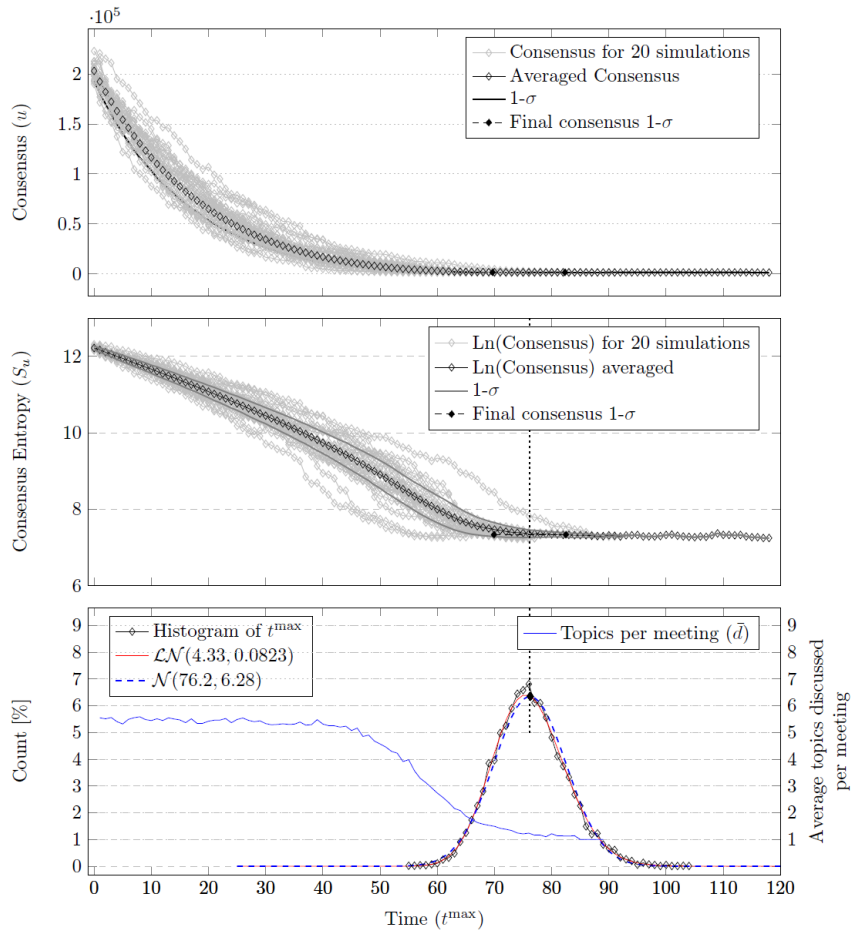
### **1.6 This Paper**

In earlier papers we developed mathematical and simulation models based on stochastic processes in a MAS simulation to study consensus and subversive agent behaviour (Vorster and Leenen, 2023a; Vorster and Leenen 2023b). In this paper we aim to explain the behaviour of the models without using mathematics and stochastic terminology, but rather, focus on the implications of the results from the simulations and in particular the application within the IO and PsyOps environments, although some results can also apply to autonomous consensus, such as for self-driving cars and UAVs.

## **2. Characteristics of Consensus**

In this section we discuss the general approach to modelling consensus using a multi-agent simulation system. First, we start with the consensus process without the presence of subversive agents. This is done to act as a baseline against which the behaviour of subversive agents can be juxtaposed.

We constructed a simulated multi-agent environment where agents must reach consensus on many topics and each agent starts the simulation with a stochastic distribution of views. Agents meet in pairs and discuss a random number of topics, from 1 to 10, with an average of 5.5 topics discussed per meeting. Agents will keep track of the topics and their view of each topic and understand if they differ in view on a topic from another agent. By comparing the view of agents pairwise with each other and summing the differences it becomes possible to construct a measure of the group consensus. Figure 1 (top graph) shows the trajectory of this consensus measure over time.



**Figure 1: Measures of Consensus**

Like earlier findings, the total consensus measure decreases exponentially (e.g. Wie, 2021). However, something that has emerged in our simulations that were not part of earlier consensus approaches is that real-world individuals need to meet and discuss topics, that is, spend time discussing matters. In the traditional MAS simulations, the agents communicate their views on all topics simultaneously to other agents. This is not realistic in a real-world scenario where it takes time to discuss topics and a given meeting has a limitation on the number of topics that can be discussed. The consequence of the limitations on number of topics that can be discussed can be seen when plotting this exponential decrease in consensus on a log-scale graph, Figure 1 (middle). If the decrease was exponential throughout the process, then the log-scale graph would remain linear, however, as can be seen in the graph, the consensus process is greatly affected by the number of topics available for discussion. Once two individuals have less than ten open topics for discussion, the meetings become much less efficient, and the consensus measure decreases more slowly.

The consequences of this finding for real-world consensus processes are that it may seem that the individuals are reaching consensus quickly on several topics, but the last few topics can take a significant time to resolve. Projects aiming to reach consensus on many topics involving many individuals may see early process in consensus measures but will quickly find a plateau of slow progress as many individuals must speak to many other individuals, but, only on a small number of topics. If one individual has 10 open items to discuss but with ten different individuals, it will take ten meetings, which is highly inefficient compared to when those ten items was related to only one or two other individuals and thus will need only a small number of meetings to resolve. The blue graph in Figure 1 (bottom) shows the evolution of the number of topics that are discussed per meeting. Initially the average is  $5.5 = (10+1)/2$  because a meeting will always discuss at least one topic and a maximum of ten topics. However, as time progresses, this number decreases and towards the end of the simulation meetings are dominated by one-topic discussions.

### 3. Subversive Agents

After a rather lengthy introduction, necessitated by the novelty of the topic, we now turn to subversive agents and their behavioural models.

Unlike the individuals with *committed views* discussed in the context of social influence models earlier, the aim of subversive agents is not to spread a specific view, but rather to spread views to increase the diversity of views, to promote views in a way that would lead to less consensus among individuals, thus delaying consensus formation on a group level. These agents may promote *view-a* to one individual and *view-b* to another individual, if that would lead to the two individual having views that evolve apart from each other.

Subversive agents also do not take extreme positions since that could lead to their exposure. They promote views already present within the group. That is, if it is known that some member holds *view-a*, then a subversive agent can promote that view to another member of the group. As the group's views evolve and contract, so do the positions promoted by the subversive agents.

#### 3.1 Subversive Agent Behaviour

We quickly discuss the behaviour of subversive agents within the group. Since agents meet in pairs to discuss outstanding topics (topics where group consensus has not been reached) the subversive agents who understand the position of their counterpart can take up a different position and try to convince the counterpart of their view. Subversive agents select the view they choose to portray from the diversity of the group views in such a way that it falls outside the central views, but not so far away that they will seem to have radical views. Figure 4 below shows these three potential positions, and our subversive agents take up positions within the barred blue area illustrated in the figure. We investigated a number of distribution, but the results reported here is not dependent on distribution of views used, and so we only report on the findings using a Normal distribution to reduce complexity of reporting.

#### 3.2 Small Numbers of Subversive Agents can Critically Harm Project Delivery

Now that the background context for subversive agents has been created, we turn to answer some questions about the effectiveness of such subversive agents. The first question that we want to answer is: can a single subversive agent have any real effect on project delivery?

Figure 2 shows the consensus trajectories for a group of twenty individuals among which are subversive agents. This is a graphical view of the effect that subversive agents have on the consensus process. As can be seen from the shift in slope of the graphs, subversive agents cause the consensus process to contract slower, and thus take much longer to complete.

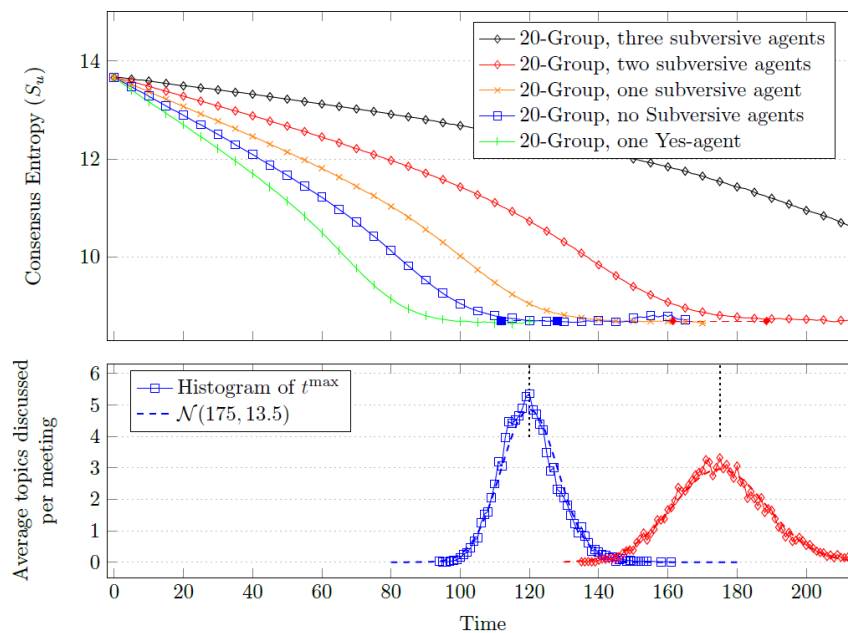


Figure 2: Impact of subversive agents on the time to reach consensus for a small number of contentious issues

If the project requires twenty individuals to complete, and it would have taken 120 time units to complete (hours, days, weeks), then the introduction of a single subversive agents will, on average, delay the project to take 139 time units (on average) to complete, which translates into a 16% project time delay, and thus also an increase of 16% in project costs. If two subversive agents are present, 10% of the group size, then they can cause a 48% increase in project costs and time, according to our simulations. Three subversive agents can more than double the project cost and time (127%), see Table 1.

**Table 1: Subversive agents can extend the time to reach consensus significantly. This can be translated into project delays. The table shows the delays that subversive agents can cause on a project consisting of twenty individuals**

Subversive Agents	Subversive Agents/Group %	Time to reach consensus	% Project delay (1 $\sigma$ )
0	0%	120	0% (14%)
1	5%	139	16% (14%)
2	10%	175	48% (16%)
3	15%	272	127% (22%)
4	20%	568	373% (42%)

### 3.3 Subversive Agent Cooperation

Since more than one subversive agent can be operationally active within the same project, one can ask if there are modes of cooperation between the agents that would improve their performance, as measured by their ability to delay the project?

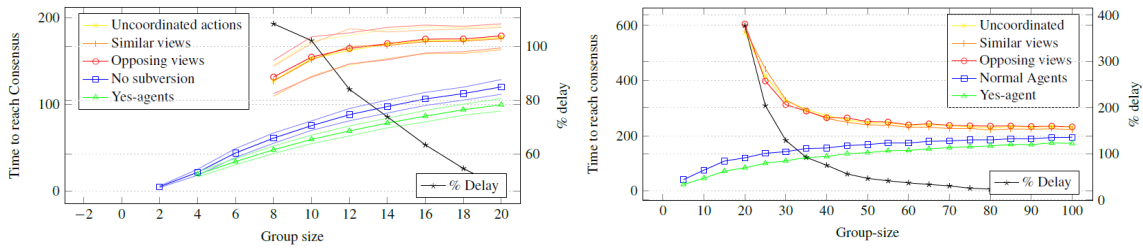
Let us first consider ways in which these agents can cooperate. Firstly, (a) they can just not cooperate at all, that is, they act using their own information without being affected of by affecting other subversive agents. Intuition would dictate that this should be the least efficient way that agents can cooperate. There are at least two trivial other ways to cooperate, namely (b) for the agents to take up the same positions trying to sway more of the group to that position; and (c) for the agents to take up opposing views to attempt split the group into two factions.

Some of the results from simulations on these conditions are given in Table 2 below. There are too many datapoints to list, so we give a small representation in Table 2, and graph all the results in Figure 3.

The data from Table 2 suggests that all three these coordination types seem to have the same or close to the same effect. For two agents in a group of 20 using (a) no cooperation, (b) cooperation to take the same position, and (c) cooperation to take opposing positions, give delays of (a) 47.5%, (b) 46.2%, and (c) 49.0% respectively, each with a standard deviation of slightly over 15%. These values are so close that the maximum difference between them (3.8%) is well within the  $\sigma=15\%$  boundary. These values are so close together that they look like a single point on the leftmost graph of Figure 3 (group-size 20 reading).

**Table 2: Subversive agent coordination types and the effect that has on project delays**

SAs/Group	Agent coordination	% delay (1 $\sigma$ )
2/20	No cooperation	47.5 (15.3)
	Same position	46.2 (15.6)
	Opposite position	49.0 (15.5)
4/40	No cooperation	75.3 (12.4)
	Same position	68.6 (12.1)
	Opposite position	70.4 (13.9)

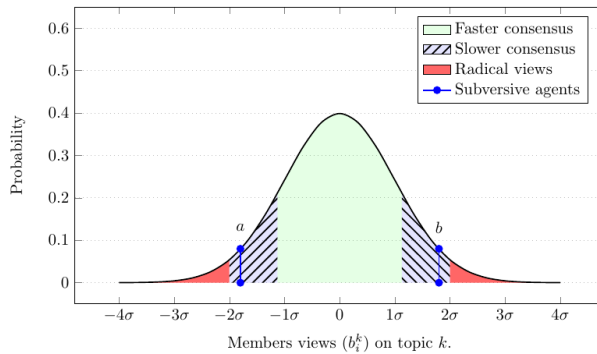


**Figure 3: The graphs show the effect of cooperation between subversive agents. Three cooperative strategies are investigated, (a) uncoordinated agents, (b) same-view agents, (c) opposing-view agents. (Left) Two subversive agents operate within various group sizes. (Right) Four subversive agents operating**

This is not an exceptional case and by visual inspection of Figure 3 (left) for group sizes 8 to 20 with two subversive agents and Figure 3 (right) for group sizes 20 to 100 with four subversive agents. This result is counter intuitive and warrants further analysis and discussion.

Let us assume that the views of all the individuals on a specific topic has a Normal distribution. Literature on this topic suggest that the distribution is highly dependent on the problem itself, and various distributions have been observed in the wild, from Normal, Uniform, Skew, Exponential, and even Double-normal (Den Boon and Van Meurs, 1991; Lang et al 2018). Our pic of distribution is illustrative and the argument we make will be valid for other distribution as well.

Any agent that takes a position within the centre 50% of the distribution of views, will thus promote a centrist view to the other 50% (see the central green area in Figure 4). If an agent takes up a view on the outside of this central region, that is the blue barred region or the outlying red region, then their efforts will increase the time to reach consensus. From the data we speculate that what is happening is that by taking positions in the outlying regions, the subversive agents widen the distribution, and it is this widening that is causing slower consensus formation. That is, it does not matter if agents expand the distribution of views to the left, or to the right, in both cases they are widening the distribution of views itself.



**Figure 4: A theoretical distribution of views on a specific topic. If a subversive agent promotes a view in the central region (solid light green) then they are promoting consensus views, and it will lead to faster consolidation of views. If a subversive agent promotes a view**

In Vorster and Leenen (2023b) we proposed a partial mathematical model to explain the results that agent cooperation does not affect the delay, but that model is not complete and thus some more work is needed to fully understand the observed behaviour. What the model and the data seem to imply is that subversive agents have the same effect no matter what their mode of working together is. It must be said that the irrelevance of cooperative model holds for the three modes we investigated. Maybe there are more sophisticated ways to cooperate that we did not investigate, that will yield better results.

For future work we propose the following additional cooperative models:

- Cooperate on specific topics, rather than on random topics. That is, maybe subversive agents will be more successful if they ignore topics with low diversity in views, and they should rather focus on specific topics with high diversity in views.

- Identify specific individuals within the group that have a high number of non-consolidated topics and focus to influence them specifically rather than targeting random members of the group.
- In our model topics are independent of each other, maybe if there were dependencies between topics subversive agents will find it easy to exploit one topic to create dependency-chaos in another topic.

#### 4. Conclusion

In this paper we explored the potential effects that subversive agents can have on large projects such as is typically found in the military industrial complex.

Subversive agents have two goals, firstly, to extend the project life through delays in consensus formation within these large project teams, and secondly to not have such extreme views that they are seen as being difficult or sowing discord. They do this by promoting views that fall outside the central set of views, but too far outside. Subversive agents adapt their views as the group's views evolve, so that they eventually also agree to the consensus view.

We found that subversive agents can have a significant impact on the time to reach consensus in these projects, and a small number of subversive agents, 10% of the group, can easily delay a project by 45 to 50% for smaller groups, and by 70 to 75% for larger groups.

Furthermore, it seems from these simple models that the mode of cooperation between subversive agents do not have a significant impact on the delays they can achieve. This is an interesting result from an espionage perspective, because it implies that such agents or their handlers do not have to risk discovery through coordinated actions, uncoordinated actions will have the same effect.

#### References

- Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., Abu-Salem, H. (2009). *A taxonomy of an IT project failure: root causes*. International Management Review, 5 (1), pp 93-116.
- Bao, G., Ma, L., Yi, X. (2022). *Recent advances on cooperative control of heterogeneous multi-agent systems subject to constraints: A survey*. Systems Science & Control Engineering, 10 (1), pp 539-551.
- Bass, F.M. (1969). *A new product growth for model consumer durables*. Management science, 15 (5), pp 215-227.
- Charchian, Daniel J. (2001). *Understanding Culture and Consensus Building: Requisite Competencies for Interagency Operations*. US Army War College, Carlisle Barracks, PA, USA. Online: <https://apps.dtic.mil/sti/pdfs/ADA390805.pdf>, accessed 2024/01/22.
- Cheung, S.Y., Gong, Y., Wang, M., Zhou, L., Shi, J. (2016). *When and how does functional diversity influence team innovation? the mediating role of knowledge sharing and the moderation role of affect-based trust in a team*. Human relations, 69 (7), pp 1507-1531.
- Den Boon, A. K. and Van Meurs, A. (1991). *Measuring opinion distributions: An instrument for the measurement of perceived opinion distributions*. Quality and Quantity, 25(4), pp 359-379.
- Enyinda, C.I., Blankson, C., Cao, G., Enyinda, I.E. (2022). *Why cannot we all just get along? resolving customer-focused team interface conflicts in a B2B firm leveraging AHP-based multi-criteria decision-making*. Journal of Business & Industrial Marketing, , <https://doi.org/10.1108/jbim-02-2021-0104>.
- Granovetter, M. (1978). *Threshold models of collective behavior*. American journal of sociology, 83 (6), 1420-1443.
- Floyd, Steven W., and Bill Wooldridge. (1992). *Managing strategic consensus: the foundation of effective implementation*. Academy of Management Perspectives, 6 (4), pp 27-39.
- Hirvonen, P. (2019). *Positioning, conflict, and dialogue in management teams*. Qualitative Research in Organizations and Management: An International Journal, 14 (4), pp 444-464.
- Kian, M.E., Sun, M., Bosché, F. (2016). *A consistency-checking consensus-building method to assess complexity of energy megaprojects*. Procedia-social and behavioral sciences, 226 , pp 43-50.
- Lang, J. W., Bliese, P. D., and de Voogt, A. (2018). *Modeling consensus emergence in groups using longitudinal multilevel methods*. Personnel Psychology, 71(2), pp 255-281
- Maldonado, M. M. (2015). *Qualitative Case Study on F-35 Fighter Production Delays Affecting National Security Guidance*. Doctoral dissertation, Walden University.
- McDonald, B. N. (1995). *Evaluating foreign-source dependencies in the US Army's M1 Abrams tank*. Doctoral dissertation, Monterey, California. Naval Postgraduate School.
- Paletz, S.B., Chan, J., Schunn, C.D. (2017). *The dynamics of micro-conflicts and uncertainty in successful and unsuccessful design teams*. Design Studies, 50 , pp 39-69.
- Vorster, J., & Leenen, L. (2023a). *Consensus simulator for organisational structures*. Proceedings of the 13th international conference on simulation and modelling methodologies, technologies, and applications (pp 15-26).

- Vorster, J., & Leenen, L. (2023b). *Exploring the effects of subversive agents on consensus-seeking processes using a multi-agent simulator*. Proceedings of the 13th international conference on simulation and modelling methodologies, technologies, and applications (pp 104–114).
- Waheeb, R.A., & Andersen, B.S. (2022). *Causes of problems in post-disaster emergency re-construction projects—Iraq as a case study*. Public Works Management & Policy, 27 (1), pp 61–97.
- Wei, Qinglai, Xin Wang, Xiangnan Zhong, & Naiqi Wu (2021). *Consensus control of leaderfollowing multi-agent systems in directed topology with heterogeneous disturbances*. In: IEEE/CAA Journal of Automatica Sinica 8.2, pp 423–43.
- Whitney, K.M., & Daniels, C.B. (2013). *The root cause of failure in complex it projects: Complexity itself*. Procedia Computer Science, vol 20 , pp 325–330.
- Xie, J., Sreenivasan, S., Korniss, G., Zhang, W., Lim, C., Szymanski, B.K. (2011). Social consensus through the influence of committed minorities. Physical Review E, 84 (1), 011130.
- Yang, R., Liu, L., Feng, G. (2022). *An overview of recent advances in distributed coordination of multi-agent systems*. Unmanned Systems, 10 (03), pp 307–325.