Private Blotto: Viewpoint Competition with Polarized Agents

Kate Donahue[1], Jon Kleinberg[1,2],

[1]Department of Computer Science, Cornell University
[2]Department of Information Science, Cornell University
kdonahue@cs.cornell.edu, kleinberg@cornell.edu

Abstract

Social media platforms are responsible for collecting and disseminating vast quantities of content. Recently, however, they have also begun enlisting users in helping annotate this content - for example, to provide context or label disinformation. However, users may act strategically, sometimes reflecting biases (e.g. political) about the "right" label. How can social media platforms design their systems to use human time most efficiently? Historically, competition over multiple items has been explored in the Colonel Blotto game setting. However, they were originally designed to model two centrally-controlled armies competing over zero-sum "items", a specific scenario with limited modern-day application. In this work, we propose and study the Private Blotto game, a variant with the key difference that individual agents act independently, without being coordinated by a central "Colonel". We completely characterize the Nash stability of this game and how this impacts the amount of "misallocated effort" of users on unimportant items. We show that the outcome function (aggregating multiple labels on a single item) has a critical impact, and specifically contrast a majority rule outcome (the median) as compared to a smoother outcome function (mean). In general, for median outcomes we show that instances without stable arrangements only occur for relatively few numbers of agents, but stable arrangements may have very high levels of misallocated effort. For mean outcome functions, we show that unstable arrangements can occur even for arbitrarily large numbers of agents, but when stable arrangements exist, they always have low misallocated effort. We conclude by discussing implications our results have for motivating examples in social media platforms and political competition.

1 Introduction

Over the last several decades, social media platforms have become hubs of information, responsible for collecting and disseminating vast quantities of content. The sheer scale of content means that traditional sources of annotation and curation (e.g. traditional fact-checking) has become borderline infeasible, even as these platforms have become primary sources of information for many people (e.g. see (Nielsen and Schrøder 2014; Gil de Zúñiga, Jung, and Valenzuela 2012)). Instead, some of these platforms have turned to other solutions, including using the platforms and users themselves to curate and annotate content. For example, the Community Notes tool on X.com (formerly known as the Birdwatch tool on Twitter (Wojcik et al. 2022)) has the goal of identifying misinformation by allowing X users to vote on "notes" with added context that are used to annotate posts.

In this work, we will draw on game-theoretic tools to help analyze scenarios like this, where multiple strategic, biased agents compete over several different items[1]. In particular, over a century ago, before modern game theory was fully established, Émile Borel proposed a family of related zero-sum games that study a similar problem to ours, in the *centralized* setting. Specifically, this models centrally-coordinated competition using military conflict as a metaphor (Borel 1921):

Definition 1 (Colonel Blotto). *Two players, A and B, are competing over M different* fronts*, with N_a, N_b units of effort at their disposal respectively. A player wins a front if they allocate more effort to the front than their opponent does, and each player wishes to win as many fronts as possible. Are there Nash stable arrangements of effort over fronts, and if so, which are they?*

The name "Colonel Blotto" comes from the fact that a colonel controls multiple individual soldiers, which they allocate across the battlefields in order to serve their overall objective. The Colonel Blotto game has been the focus of extensive exploration, including variants that allow for battlefields to have different values, for effort to be allocated probabalistically, and for smoother utility functions (Golman and Page 2009; Hart 2008; Osorio 2013) (see Appendix A.1 for more works). One common thread has been the relative scarcity of pure Nash equilibria, which has centered the literature around exploration of mixed Nash equilibria. The game has also found many applications in areas far removed from warfare (Merolla, Munger, and Tofias 2005), such as national politics, but always with centralized entities competing over multiple items.

Modeling decentralized conflicts. However, the Colonel Blotto framing is at odds with the modern type of disaggregated competition, such as our motivating example of social

---

[1]Other settings where this occurs include political competition over multiple issues: see Section 1.1 for a discussion.

media users labeling items where multiple agents may share similar goals, but are not controlled by a central organizing "Colonel". How might we model this type of political *viewpoint competition*?

We could imagine that there is a large collection of agents (e.g. users), each of whom is interested in taking part in a conflict with M "items" (e.g. online posts) Each agent controls only one unit of effort, and can choose to devote that effort to one of the items (i.e. labeling a post as misinformation or not). There is no centralized "colonel" to direct the agents, but instead each agent i has one of two *types* (represented by real numbers β_a, β_b), which we can think of as a viewpoint, bias, or political position. After each agent chooses an item to engage with, the outcome of the conflict on each given item is determined by an *outcome function* that takes the multiset of types at that item and determines a real-valued outcome. We will also include a positive penalty c for leaving an item unlabeled (empty).

An "outcome function" determines how inputs from multiple users results in a single label for the item. By selecting one outcome function over another, the designer could influence how individual agents choose to exert their effort. Thus, the choice of outcome function will be one of the central focus points of this paper – it represents one of the few aspects that the designer of the system may have control over (where the designer could be a social media company or political entity running the election, for example).

We will be particularly interested in two natural outcome functions for aggregating viewpoints in this setting: the *median* (in which the outcome on an item is the median of the types there) and the *mean* (in which the outcome is the mean of the types). Agents want the outcomes on each item (even the ones where they don't participate) to match their types; thus, each agent experiences a cost equal to the average of the distances between the outcome on each item and the agent's type.

We will refer to this type of game as *Private Blotto*; like Colonel Blotto, it involves conflict over multiple items, but it is fundamentally different because it is designed to model decentralized conflict where each individual agent makes their own choice about which item to participate in[2]. We summarize the discussion above in the following definition.

Definition 2 (Private Blotto). *Two types of agents, type A and B, are competing over M different items, with N_a agents of type A and N_b agents of type B. Each agent chooses exactly one item to compete in (label), and an outcome function (for example, the median or mean) determines the outcome value on each item. An agent's cost is equal to the average distance between the outcome on each item and the agent's type. Are there Nash stable arrangements of agents over items, and if so, what do they look like?*

For this class of games, we can explore a number of questions. One of the most fundamental contrasts we will show is that the decentralized *Private* Blotto game admits

pure Nash equilibria more frequently than the centralized *Colonel* Blotto game: part of our work's contribution will be to characterize when these equilibria occur, and what they look like. Additionally, one vital question is how the choice of outcome function affects the existence of stable arrangements, and how those stable arrangements distribute agents across items. Finally, we will explore how stable arrangements compare according to normative goals of utilizing user time well (minimizing *misallocated effort* of users onto unimportant items). We now provide some more detail on settings that can be modeled by the Private Blotto game, and then we give an overview of our results.

1.1 Motivating Examples and Further Related Work

Our Private Blotto formulation finds applicability in numerous modern-day settings. Here, we will describe a few key application areas in more detail.

Crowdsourcing on social media has become a growing area of societal and academic interest in recent years (Yasseri and Menczer 2021; Wojcik et al. 2022; Allen, Martel, and Rand 2022; Pröllochs 2022). Focusing on Community Notes, users can provide (discrete) labels on tweets, labels which have been shown (Allen, Martel, and Rand 2022) to have partisan bias. Empirical studies (Allen, Martel, and Rand 2022; Saeed et al. 2022) show most tweets only have 1-2 labels and the modal user submits only one note (median user submits 5), which matches the Private Blotto setting of discrete labels with bandwidth-limited agents. Other aspects of Community Notes match Private Blotto well: users are given pseudonyms when voting, making coordination between users implausible, and typically see the labels on each tweet before deciding to label it, mirroring how users in Private Blotto determine how to allocate effort based on the existing set of labels (Wojcik et al. 2022). Because Community Notes users are also X users, they view tweets that aren't already labeled, and may incur some disutility for leaving misinformation unlabeled, motivating Private Blotto's positive cost c for leaving an item empty. We describe other related papers in crowdsourcing in Appendix A.2.

Separately, our Private Blotto setting also finds applications in political contests or issue-based activism, as well as military engagements, which have both historically been application areas for Colonel Blotto (Merolla, Munger, and Tofias 2005). For political contests, we argue that Private Blotto might even be a more natural fit. Here, the M items might represent issues or political campaigns, while the agents might be activist groups or donors, which might share similar goals, but are unable (for logistical or legal reasons) to coordinate their actions. Differing types would reflect differing political leanings, which could be closer or further apart (reflecting the magnitude of the gap in biases). For military engagements, in Private Blotto each agent might be an individual soldier, guerrilla member, or other actor that is acting without coordination from some central organizer. In this way, Private Blotto might naturally model more modern types of asymmetric warfare conflicts, where agents on the same "side" militarily are of the same type.

---

[2]In the military, a *private* is an enlisted soldier at the base of the hierarchy. This reflects our setting, which views the individual soldiers as the strategic actors, rather than the coordinating colonel who commands the army.

1.2 Overview of Results

We are primarily interested in which instances of Private Blotto games admit *stable arrangements* where no agents has an incentive to unilaterally change items. We are also curious about the properties of stable arrangements, when they exist – do they result in agents being spread out across items, or could they involve many agents clustered on a single item?

We will find it useful to divide our analysis of the model into two main cases, depending on whether there are more agents than items (Section 3) or more items than agents (Section 4).

When there are more agents (Section 3), we will find that the choice of outcome function (mean or median) can produce starkly different results. We can view these results as considering the plane of (N_a, N_b) pairs and asking whether a stable arrangement must exist for N_a type A players and N_b type B players. In the case of the median outcome function, we show that there is a *median-critical region* of bounded size where stable arrangements fail to exist: in particular, this means that given sufficiently many agents, a stable arrangement is always guaranteed to exist. However, these stable arrangements could have almost all agents clustered on a single item. If we take the normative principle that items (representing posts or political issues) should all receive approximately equal levels of attention, such disproportionate levels of agents could be viewed as a high level of *misallocated effort* on the part of agents.

By contrast, the mean outcome function results in very different results for stable arrangements. In particular, we show that there are arbitrarily many (and arbitrarily large) (N_a, N_b) pairs where no stable arrangement exists, showing that stability may be much harder to guarantee. However, we also show that when a stable arrangement exists, it must have all agents split almost exactly evenly across items (up to integer rounding), resulting in a very low level of misallocated effort.

In Section 4 we turn to the scenario where there are more items than agents: intuitively, these are settings where some items will inevitably be left empty. Here, we show that median and mean outcome functions produce very similar outcomes, which is natural given that the mean and median are identical for small numbers of items. At a high level, while we show that while settings without any stable arrangements can frequently exist, so long as the cost for leaving an item empty is sufficiently high, there is always a stable arrangement where players spread out with exactly one player per item.

Finally, Section 5 concludes and discusses implications that our results may have for our motivating examples in post annotation in social media and political competition over issues. In particular, our results show that the choice of outcome function can dramatically influence how biased agents may choose to expend their effort.

2 Model

In this section, we make our theoretical model more precise. We assume there are M *items* that N total *agents* are competing over. Each agent controls exactly 1 unit of effort: they may choose which item to compete in, but may not coordinate with other players. However, agents come in two *types* (A and B). Two agents of the same type have perfectly aligned incentives: when present on the same item, they work towards the same outcome, and when on different items, two agents of the same type are interchangeable. Each type has a real-valued *bias* $\beta_a, \beta_b \in \mathbb{R}$ that describes how similar or dissimilar the types are to each other (how polarized the two groups are). For example, $\beta_a = 1, \beta_b = -1$ agents are closer to each other than $\beta_a = 5, \beta_b = -3$.

2.1 Outcome Functions

Once agents are arrayed on an item, the outcome of the battle is governed by an *outcome function* $f(\cdot)$. In this paper, we will focus on two types of outcome functions: *median* outcome and *mean* (proportional) outcome. Given a set of agents S_i on item i, the median outcome function returns the median of the biases $\text{med}(\{\beta_t\} \mid t \in S_i)$. If there are an even number of players on a particular item, then the median function averages together the middle two biases. Given two types of players, the median outcome function is equivalent to "winner-take-all", where whichever type dominates the item wins.

On the other hand, the mean outcome function returns the mean of the biases: $\frac{1}{|S_i|} \sum_{t \in S_i} \beta_t$. This models a scenario where the final outcome of the item depends on the distribution of agent biases, not solely the median agent.

For both mean and median outcome function, we assume agents have cost given by the distance of their bias to the outcome: that is, given outcome function $f(\cdot)$ on an item with set of labels S_i, then an agent with bias β_t has cost for that item of $|f(S_i) - \beta_t|$.

2.2 Agent Cost

Even though each agent only participates in a single item, we model the agents as still having preferences over all of the items. This could reflect settings where the agents are social media users who observe multiple posts but only have the bandwidth to provide annotations on a smaller subset, or political actors who have opinions about many topics but only focus their energy on a single issue. Additionally, we assume each agent experiences a cost $c \geq 0$ for leaving the item empty, which is independent of agent bias. We include this feature to model settings where agents may choose to leave an item empty (to join a more contested item), but suffer some non-zero cost in doing so. We can write the total cost as:

$$\sum_{i \in [M], |S_i| > 0} |f(S_i) - \beta_t| + \sum_{i \in [M], |S_i| = 0} c$$

We will say that an arrangement of agents across items is *stable* if it satisfies Nash stability (no agent can unilaterally decrease its cost). In the online content labeling setting, if an arrangement fails to be Nash stable, this means that some online users would prefer to move which posts they label, or potentially generate more labels — which could cause unstable cycles where users compete in an arms race to generate more posts.

Definition 3 (Nash stability (pure)). *An arrangement of players on items is (Nash) stable in the Private Blotto game if no agent can reduce its cost by switching from competing in one item to begin competing in another.*

Inherent in this definition of stability is the requirement that agents are *decentralized*: in particular, each agent is deciding which action to take by themselves, without coordinating among other agents of the same (or similar) biases. This definition is a departure from the prior Colonel Blotto literature which allowed multiple agents to be coordinated by a central "Colonel". However, Definition 3 is the natural definition of stability to study for more decentralized settings (e.g. crowdsourcing, disaggregated political contests) with self-interested actors. Note that we focus on the most natural question of *pure* Nash equilibria, although exploring mixed Nash equilibria would be an interesting extension for future work.

3 More Agents: $N \geq M$

In analyzing the Private Blotto game, we will find it helpful to divide our analysis into two main regimes: when there are more agents than items (this section), and when there are fewer agents than items (Section 4). As related to the examples in Section 1.1, in online crowdsourcing this models scenarios where there are enough online users that they could provide each post with at least one label, and in political issues it could represent the case where there is a relatively small subset of major divisive issues that multiple political actors are debating. Without loss of generality, we will always name the two types so that $N_a \geq N_b$ (there are more type A than type B players).

Lemma 1 shows the setting with more agents than items is especially clean: so long as the cost for leaving an item empty is sufficiently high, then no item will be left empty, and all results will be completely independent of agent biases β_i.

Lemma 1. *If there are more agents than items ($N \geq M$) and the cost for leaving a item empty is sufficiently high, then no item will be left empty, regardless of if median or mean outcome function is used. Specifically, this occurs when:*

$$c \geq 0.5 \cdot |\beta_a - \beta_b|$$

Moreover, agent strategy becomes independent of biases β_a, β_b and relies solely on the number of agents of each type on each item, $\{a_i, b_i\}, \ i \in [M]$.

Proofs for Lemma 1, as well as for other proofs in this paper, are found in Appendix C. For the rest of this section, unless stated otherwise, we will assume that the preconditions of Lemma 1 hold, which ensures that no item will be left empty. This assumption is mainly made for cleanness of analysis: if it is relaxed, then the value of c causes minor changes in the stable arrangements, primarily for small numbers of agents N.

3.1 Median Outcome

First, in this section we explore Private Blotto with $N \geq M$ with median outcome function, modeling the case where contests are decided by a winner-take-all outcome. We can view this setting as exploring the N_a, N_b plane, studying for which values of N_a, N_b a Nash equilibrium exists, as well as constructively producing an example of a stable arrangement. Our results will be a function of the total number of items (M), as well as N_a, N_b, the number of agents of types A and B respectively. (Recall that given Lemma 1 all of our results will be independent of the biases β_a, β_b.)

Our main result for this section is Theorem 1, which exactly characterizes when a stable arrangement exists for median outcome. Specifically, this occurs whenever the number of agents N_a, N_b for each of the types is *not* in the median-critical region (Definition 4).

Definition 4 (Median-critical region). *A set of parameters (N_a, N_b) is in the median-critical region if they satisfy:*

$$N_a + N_b \leq 2 \cdot M \text{ and } M < N_a \text{ and } 1 \leq N_b < N_a - M$$

and symmetrically if the roles of N_a, N_b are reversed.

Theorem 1. *Given more agents than items ($N_a + N_b \geq M$), there exists a stable arrangement if and only if (N_a, N_b) is not in the median-critical region.*

We will prove Theorem 1 through several sub-lemmas which collectively handle different values of (N_a, N_b). First, Lemma 2 proves that any set of parameters within the median-critical region must always result in an unstable arrangement.

Lemma 2. *Given median outcome and cost satisfying Lemma 1, for any set of biases β_a, β_b, for all instances within the median-critical region, there is never a stable arrangement of agents onto items.*

Next, we will prove that any number of agents (N_a, N_b) outside of the median-critical region must have a stable arrangement. Lemma 3 handles the case where there are more than twice as many agents as there are items, showing that this implies there must always exist a stable arrangement.

Lemma 3. *Given median outcome and cost satisfying Lemma 1, if $N_a + N_b \geq 2 \cdot M + 1$ (or $N_a + N_b = 2 \cdot M$ with N_a, N_b even), then there always exists a stable arrangement.*

Finally, we address the question of (N_a, N_b) pairs with $N_a + N_b \leq 2 \cdot M$ (so that they are not addressed by Lemma 3), but which also do not fall in the median-critical region (so they are not addressed by Lemma 2). Lemma 4 examines this case and constructively shows that it is always possible to find a stable arrangement of agents onto items.

Lemma 4. *Given median outcome and cost satisfying Lemma 1, any number of agents (N_a, N_b) with $N_a + N_b \leq 2 \cdot M$ (besides those in the median-critical region) always has a stable arrangement.*

Taken together, these lemmas prove Theorem 1, exactly characterizing when a stable arrangement exists in Private Blotto games. In particular, they show that unstable arrangements are relatively *rare*, constrained only to the median-critical region (Definition 4). For small M, this region can be very small. For example, it is empty for $M = 2$, implying that there is always a stable arrangement for median
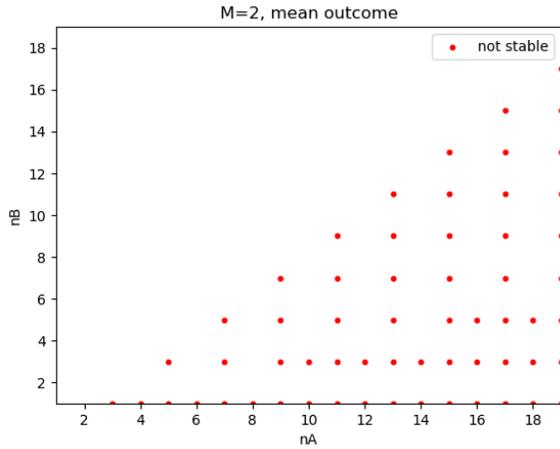
Figure 1: Figure illustrating when stable arrangements of agents onto items fail to exist for mean outcome ($M = 2$ items). For clarity, only displayed for $N_a \geq N_b$.

outcome function with two items. For $M = 3$, the median-critical region contains only two points ($N_a = 4, N_b = 1$), ($N_a = 5, N_b = 1$), showing that for almost all (N_a, N_b) pairs, a stable arrangement must exist.

3.2 Mean Outcome

Next, in this section we will explore the setting where again there are more agents than items ($N \geq M$), but where instead the mean outcome function is used. At a high level, in Section 3.1 we proved that there were only a finite number of pairs (N_a, N_b) such that no stable arrangement of players onto items existed. By contrast, in this section we will show that for mean outcome function, there are arbitrarily many pairs (N_a, N_b) with no stable arrangement.

For illustrative purposes, Figure 1 numerically explores when a stable arrangement exists for $M = 2$ items[3]. The axes represent the total number of type A and type B agents that are present, with a red dot appearing at point (N_a, N_b) if no possible stable arrangement involving that number of agents exists. Note that the red dots extend for high values of N_a, N_b, indicating that stable arrangements can fail to exist even for large numbers of players. This is in stark contrast to the median function in Section 3.1, where we showed that a stable arrangement only fails to exist within the (bounded) median-critical region. For example, the corresponding plot to Figure 1 for median would not have any red dots at all (given that the median-critical region is empty for $M = 2$ items).

Theorem 2 formalizes the intuition from Figure 1, proving that for any arbitrarily large set of players N_a, N_b, it is possible to find $N_a' \geq N_a, N_b' \geq N_b$ such that no stable arrangement of players onto items exists when mean outcome is used. In particular, the construction within Theorem 2 involves $N_a = N_b + 2$, with both N_a, N_b odd: note that in

Figure 1, all such points (odd numbers exactly 2 apart) have red dots, indicating no stable arrangement exists.

Theorem 2. *For every $N_a \geq N_b$, there exists an $N_a' \geq N_a, N_b' \geq N_b$ and M such that there is no stable arrangement of N_a', N_b' players onto M items.*

What is driving this persistent pattern of instability? Consider a hypothetical variant of Private Blotto where agents can allocate effort fractionally across items[4]. Theorem 3 shows that this would always lead to a stable arrangement exactly equal to even allocation over items. Thus, persistent recurrence of instability in Figure 1 are driven solely by the requirement that effort be allocated in integer units.

Theorem 3. *For M items with two types of agents, A and B with mean outcome and c satisfying the conditions of Lemma 1, if players are allowed to be allocated fractionally over items, then the stable arrangement is always given by $a_i = N_a/M, b_i = N_b/M$.*

Note that, in general, Theorem 3 does *not* imply that stable arrangement for the integer-valued Private Blotto games will be close to proportional. While Theorem 3 can be extended to show that the fractional Private Blotto game is convex, it is known that in general, the minimum of a convex function, when restricted to integer values may be arbitrarily far from the minimum of the same convex function over real numbers[5]. However, it turns out that for the Private Blotto game it *is* true that integer-valued stable arrangements are "close" to proportional. This idea is formalized in Theorem 4:

Theorem 4. *Given mean outcome function, any arrangement that is stable must be "close" to proportional: $|a_i - N_a/M| \leq 1, |b_i - N_b/M| \leq 1$ for $i \in [M]$, given c satisfying the conditions of Lemma 1.*

3.3 Misallocated Effort

Finally, in this section we will compare the stable arrangements given either median or mean outcome functions. In particular, we will consider the question of how "bad" stable arrangements might be, which may influence which type of outcome function might be most desirable for a given contest. This question has been formalized in a variety of ways in previous papers, including Price of Anarchy or Price of Stability (Koutsoupias and Papadimitriou 2009; Anshelevich et al. 2008). For example, in a congestion routing game, Price of Anarchy would measure the total congestion for all players in the worst-case stable arrangement, as compared to the arrangement that minimizes total congestion.

However, Private Blotto is modeling a fundamentally different game. In the plaform annotation scenario, Private Blotto is modeling different online users competing over posts for which they have truly different viewpoints, influenced by their personal biases and knowledge. In the absence of impartial, "gold standard" misinformation investigation (which may be impossible to do at scale), the social

---

[3]Code to reproduce figures and numerical examples is available at https://github.com/kpdonahue/private_blotto

[4]In this setting player payoff is similarly given by the mean outcome, but with $a_i, b_i \in \mathbb{R}$: see Appendix C for details.

[5]For an illustrative example, see (math.stackexchange.com 2015) cited in the references.

media company (and society at large) may not have a clear sense of which misinformation label is "right". However, we may have a normative preference that all posts should obtain roughly equal numbers of labels. For example, if one post has dozens of battling annotations while other posts go unlabeled, we might view that allocation of human effort as wasteful[6]. This intuition of "misallocated effort" is formalized below:

Definition 5. *Given an arrangement of agents onto items, we say it has "misallocated effort" given by the amount of agents that is above or below equal allocation. That is, misallocated effort is given by:*

$$\sum_{i \in [M]} \left| \frac{N_a}{M} - a_i \right| + \left| \frac{N_b}{M} - b_i \right|$$

One question we will explore is the maximum possible misallocated effort, among any stable arrangements. Here, we will show a qualitative difference in the bound depending on whether mean or median outcome functions are used.

First, Lemma 5 shows that misallocated effort is upper bounded by a constant times the number of items M, driven by the results of Theorem 4.

Lemma 5. *For mean outcome function, misallocated effort is upper bounded by $2 \cdot M$.*

Next, Lemma 6 shows that worst-case misallocated effort can be much higher for the median outcome function, especially in the case where there are many more agents than items. Specifically, Lemma 6 shows that misallocated effort could be as high as $N_a \cdot \left(1 - \frac{1}{M}\right)$, which can be much greater than the $2 \cdot M$ bound given in Lemma 5.

Lemma 6. *For median outcomes, worst-case misallocated effort is* lower*-bounded by $0.25 \cdot N$, given $N = N_a + N_b \geq 2 \cdot M$.*

This proof (deferred to Appendix C) is constructive and creates an arrangement with high misallocated effort, based on the proof of Lemma 4. This lower bound involves an arrangement that is a (pure) Nash equilibrium under Definition 3 because no *single* player can decrease its cost by changing which item it is occupying. However, if agents were allowed to coordinate, as in *Colonel* Blotto, then such an arrangement may fail to be stable (e.g. multiple agents of type one type could move together to another item and decrease their total cost). We view this result as illustrating how the decentralized nature of the Private Blotto game can cause arrangements with high misallocated cost to be stable, where in the centralized Colonel Blotto they would be unstable.

Overall, the results of this analysis imply that mean outcome functions, rather than median ones, give sharper guarantees that any stable arrangement that exists will result in agents roughly arranging themselves across items in proportion to their overall value. Specifically, our results have implications for the design of social media labeling tools.

---

[6]There are multiple natural extensions to the Private Blotto game, including cases where some items are more important than others, or where players come in arbitrarily many types: see Appendix B for a discussion and extension of our results.

While these tools have outcome functions that are significantly more sophisticated than median or mean (Wojcik et al. 2022), in general our results suggest that more smooth outcome functions (similar to mean) as compared to sharper outcome functions (similar to median) would reduce the worst-case misallocated effort, at the expense of greater unpredictability as to when a stable allocation exists.

4   More Items: $M > N$

In Section 3, we examined the case with more agents than items: $N \geq M$. In this section, we will explore the other possibility, with fewer agents than items. This could model content annotation in settings where the number of prospective users is far fewer than the number of posts (a common occurrence). Our goal here is to model the set of stable arrangements, again comparing median and mean outcome functions.

Because $N < M$, some items will inevitably need to be left empty. Because of this, we will drop the lower bound in Lemma 1 and allow the cost for leaving a item empty c to be set arbitrarily. Since Lemma 1 no longer applies, in this section we will see that agent biases $\{\beta_i\}$ are relevant for agents' strategies – a departure from Section 3, where all of our results held independent of agent bias.

At a high level, Section 3 showed a wide divergence in behavior between median and mean outcome functions. In this section, we will show that the setting of $N < M$ gives much more similar results between the two outcome functions, though with some differences. The intuition is that both median and mean outcome functions behave identically for items with only 1 or 2 agents present. Because there are few agents compared to the number of items, most arrangements will have 1 or 2 agents per item, unless c is very small or differences in biases is very large. Our main result in this section is Theorem 5, which exactly characterizes when a stable arrangement is guaranteed to exist and when it is possible that none exist for the $N < M$ regime. In particular, note that mean and median outcome function have almost identical patterns of when stable arrangements exist, except for at $N = 3$, and that cases where no stable arrangement exists are relatively common.

Theorem 5. *Given $M > N$,*

- *For $N = 2$, there is always a stable arrangement (for both median and mean outcome function).*
- *For $N \in [4, M)$, for both median and mean outcome function there are always parameters (player biases $\{\beta_a, \beta_b\}$ and unlabeled cost c) such that no stable arrangement exists.*
- *For $N = 3$, for mean outcome function, a stable arrangement always exists, but for median outcome function, there exists parameters such that no stable arrangement is possible.*

We prove this theorem through a series of smaller lemmas.

First, Lemma 7 considers the $N = 2$ setting, showing that a stable arrangement always exists for both median and mean outcome.

Lemma 7. *For $N = 2$, $M \geq 2$ with either median or mean outcome functions, a stable arrangement always exists, regardless of the player biases $\{\beta_a, \beta_b\}$ and unlabeled cost c.*

Next, Lemmas 8 and 9 describe the complementary condition, showing when (for median and mean outcome functions respectively), stable arrangements fail to exist. The proofs (deferred to Appendix C) are constructive and involve creating a set of agents with specific biases and cost for leaving an item unlabeled such that any possible arrangement has at least one agent wishing to label a different item.

Lemma 8. *For any N, M with such that $2 < N < M$, with median outcome, there exists biases $\{\beta_i\}$ and costs c such that no NE exists.*

Lemma 9. *For any N, M such that $4 \leq N < M$, with mean outcome, there exists parameters such that no NE exists.*

The mean outcome function case explored in Lemma 9 is not exactly analogous to the median case in Lemma 8: there is a gap at $N < 4$ players. Lemma 10 completes this gap by showing that the gap in Lemma 9 is inevitable: any possible set of $N = 3$ agents must have a stable arrangement, given mean outcomes.

Lemma 10. *For $N = 3$, $M \geq 4$ with mean outcome, there is always a stable arrangement.*

Taken together, these results prove Theorem 5. At a high level, these results show that for almost all $N < M$, it is possible that no stable arrangement of players over items exists. However, Lemma 11 shows that a stable arrangement where every agent is labeling a separate is guaranteed to exist when the cost for leaving an item unlabeled is sufficiently high:

Lemma 11. *Given $N < M$, an arrangement with all agents labeling different items is stable (for both median and mean outcome) so long as the cost for leaving an item unlabeled is sufficiently high: $c \geq 0.5 \cdot |\beta_a - \beta_b|$*

Note that the condition in Lemma 11 (when having exactly 1 agent per item is stable, given $M > N$) is exactly the same as the condition in Lemma 1 (when no item will ever be left unlabeled, given $M \leq N$). This suggests that this level of c could be viewed as a critical threshold which governs when certain types of arrangements are stable.

Taken together, these results have implications for designers of real-life systems that behave like Private Blotto games. Specifically, Theorem 5 suggests that for $N < M$, the choice of outcome function (median or mean) is relatively unimportant. However, for almost all values of N, it is possible that no stable arrangement exists, unless the cost for leaving an item unlabeled is sufficiently high (Lemma 11). A designer of such a system could increase the cost of leaving an item unlabeled by proactively highlighting posts in need of notes (in the social media example) or giving more airtime to political issues that are under-debated (in the political competition example).

5 Discussion

In this paper, we proposed and analyzed the Private Blotto game, a multi-player game involving competition over items with different values. We focused on the impact of the outcome function on whether Nash stable arrangements exist. For the case with more agents than items, we showed that the choice of outcome function is critical. A median outcome function guarantees that the number of settings where no stable arrangement exists is finite and small compared to the number of items M, but could involve high levels of misallocated effort where agents are unevenly spread across items. By contrast, the mean outcome function does not guarantee stable arrangements exist, even for arbitrarily large number of players. However, when a stable arrangement exists, it is guaranteed to have low levels of misallocated effort. In Section 4 we analyzed the case with more items than agents, showing that median and mean outcome functions behave much more similarly, and given sufficiently high cost for leaving an item unlabeled, all agents will spread evenly over items, minimizing misallocated effort.

Throughout, we used motivating examples related to civic institutions and social welfare, such as detection of online misinformation. Our results have implications for how such tools should be developed, especially in the choice of the outcome function. Specifically, if there are many agents and guaranteeing a stable arrangement is important, median outcome (or a similar function) would likely be best, but if minimizing misallocated effort is more important, mean outcome (or a similar function) is likely better. For cases with fewer agents than items, increasing the cost of leaving items unlabeled is more important than the choice of outcome function. While in this work we addressed the primary question of Nash stability, there are multiple interesting extensions into the Private Blotto game. One natural extension would be to consider cases where some items are considered more important than others - for example, we may wish to have more human effort placed on discussing an important bill than on which shoes a celebrity wore to a gala. Another natural extension would be to consider more than two polarized types A and B, potentially a continuum of biases $\{\beta_i\}$ reflecting a more nuanced set of opinions. We discuss how some of our results can be extended to both of these settings in Appendix B.

Finally, another natural extension could interpolate between the historic Colonel Blotto game and our novel Private Blotto game by allowing agents to coordinate with up to $N' > 1$ other agents. For example, for $N' = 2$, an arrangement would fail to be Nash stable if two agents, working together, could move and improve both of their utility. This modification would only make Nash stability harder to achieve, but could reflect scenarios with limited degrees of coordination.

References

Adam, L.; Horčík, R.; Kasl, T.; and Kroupa, T. 2021. Double oracle algorithm for computing equilibria in continuous games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 5070–5077.

Ahmadinejad, A.; Dehghani, S.; Hajiaghayi, M.; Lucier, B.; Mahini, H.; and Seddighin, S. 2016. From duels to battlefields: Computing equilibria of Blotto and other games. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Allen, J.; Martel, C.; and Rand, D. G. 2022. Birds of a Feather Don't Fact-Check Each Other: Partisanship and the Evaluation of News in Twitter's Birdwatch Crowdsourced Fact-Checking Program. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391573.

Anbarcı, N.; Cingiz, K.; and Ismail, M. S. 2020. Proportional resource allocation in dynamic n-player Blotto games. *arXiv preprint arXiv:2010.05087*.

Anshelevich, E.; Dasgupta, A.; Kleinberg, J.; Tardos, É.; Wexler, T.; and Roughgarden, T. 2008. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, 38(4): 1602–1623.

Behnezhad, S.; Dehghani, S.; Derakhshan, M.; Hajiaghayi, M.; and Seddighin, S. 2017. Faster and Simpler Algorithm for Optimal Strategies of Blotto Game. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

Boix-Adserà, E.; Edelman, B. L.; and Jayanti, S. 2020. The multiplayer colonel blotto game. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 47–48.

Borel, E. 1921. La théorie du jeu et les équations intégralesa noyau symétrique. *Comptes rendus de l'Académie des Sciences*, 173(1304-1308): 58.

Gil de Zúñiga, H.; Jung, N.; and Valenzuela, S. 2012. Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, 17(3): 319–336.

Golman, R.; and Page, S. E. 2009. General Blotto: games of allocative strategic mismatch. *Public Choice*, 138(3): 279–299.

Goyal, S.; and Vigier, A. 2014. Attack, Defence, and Contagion in Networks. *The Review of Economic Studies*, 81(4): 1518–1542.

Hart, S. 2008. Discrete Colonel Blotto and general lotto games. *International Journal of Game Theory*, 36(3): 441–460.

Hettiachchi, D.; Kostakos, V.; and Goncalves, J. 2022. A survey on task assignment in crowdsourcing. *ACM Computing Surveys (CSUR)*, 55(3): 1–35.

Koutsoupias, E.; and Papadimitriou, C. 2009. Worst-case equilibria. *Computer science review*, 3(2): 65–69.

Kovenock, D.; and Roberson, B. 2012. Coalitional Colonel Blotto games with application to the economics of alliances. *Journal of Public Economic Theory*, 14(4): 653–676.

math.stackexchange.com. 2015. Is the optimal solution of a strictly convex function over $\mathbb{Z}^d$ a rounded version of its optimal solution over $\mathbb{R}^d$.

Mazur, K. 2017. A Partial Solution to Continuous Blotto. *arXiv preprint arXiv:1706.08479*.

Merolla, J.; Munger, M.; and Tofias, M. 2005. In play: A commentary on strategies in the 2004 us presidential election. *Public Choice*, 123: 19–37.

Nielsen, R. K.; and Schrøder, K. C. 2014. The Relative Importance of Social Media for Accessing, Finding, and Engaging with News. *Digital Journalism*, 2(4): 472–489.

Osorio, A. 2013. The lottery Blotto game. *Economics Letters*, 120(2): 164–166.

Pröllochs, N. 2022. Community-based fact-checking on Twitter's Birdwatch platform. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 794–805.

Saeed, M.; Traub, N.; Nicolas, M.; Demartini, G.; and Papotti, P. 2022. Crowdsourced Fact-Checking at Twitter: How Does the Crowd Compare With Experts? In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 1736–1746.

Schwartz, G.; Loiseau, P.; and Sastry, S. S. 2014. The heterogeneous colonel blotto game. In *2014 7th international conference on NETwork Games, COntrol and OPtimization (NetGCoop)*, 232–238. IEEE.

Skaperdas, S. 1996. Contest success functions. *Economic theory*, 7: 283–290.

Tullock, G. 2001. *Efficient Rent Seeking*, 3–16. Boston, MA: Springer US. ISBN 978-1-4757-5055-3.

Wojcik, S.; Hilgard, S.; Judd, N.; Mocanu, D.; Ragain, S.; Hunzaker, M.; Coleman, K.; and Baxter, J. 2022. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation. *arXiv preprint arXiv:2210.15723*.

Yasseri, T.; and Menczer, F. 2021. Can the Wikipedia moderation model rescue the social marketplace of ideas?

Zhang, J.; Sheng, V. S.; Li, Q.; Wu, J.; and Wu, X. 2017. Consensus algorithms for biased labeling in crowdsourcing. *Information Sciences*, 382: 254–273.

Zhang, Y.; and van der Schaar, M. 2012. Reputation-based incentive protocols in crowdsourcing applications. In *2012 Proceedings IEEE INFOCOM*, 2140–2148.