Procedural Fairness in Multi-Agent Bandits

Extended Abstract

Joshua Caiata
University of Waterloo
Waterloo, Canada

Carter Blair
Harvard University
Cambridge, United States

Kate Larson
University of Waterloo
Waterloo, Canada

ABSTRACT

In the context of multi-agent multi-armed bandits (MA-MAB), fairness is often reduced to outcomes: maximizing welfare, reducing inequality, or balancing utilities. However, evidence in psychology, economics, and Rawlsian theory suggests that fairness is also about process and who gets a say in the decisions being made. We introduce a new fairness objective, *procedural fairness*, which provides equal decision-making power for all agents, lies in the core, and provides for proportionality in outcomes. Empirical results confirm that fairness notions based on optimizing for outcomes sacrifice equal voice and representation, while the sacrifice in outcome-based fairness objectives (like equality and utilitarianism) is minimal under procedurally fair policies. This paper argues that procedural legitimacy deserves greater focus as a fairness objective, and provides a framework for putting procedural fairness into practice.

KEYWORDS

Fairness, Multi-armed Bandits, Multi-agent Systems

ACM Reference Format:
Joshua Caiata, Carter Blair, and Kate Larson. 2026. Procedural Fairness in Multi-Agent Bandits: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/RCLN5951

1 INTRODUCTION

In the multi-agent systems we build today, fairness is almost always reduced to optimizing for a specific outcome [4–6]: the sum of utilities, the balancing of welfare, or the smoothing of inequality. However, evidence in psychology and economics shows that people consistently value fair process–even if that means a less-than-ideal outcome[1, 7, 10]. As a result, this paper begins from a new conviction: that fairness in multi-agent systems must be grounded not in optimal outcomes, but in the principle of equal voice.

What is missing in the literature is a framework that gives agents themselves an equal share of decision-making power. Inspired by Rawls' notion of *pure procedural justice* [8], we formalize procedural fairness in MA-MABs, a framework where each action (pulling an arm) produces potentially different rewards for each agent, sampled from potentially different distributions. This framework naturally captures both the allocation of benefits and the distribution of decision-making power in a simple and easy-to-understand way.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/RCLN5951

To situate procedural fairness, we compare it, both theoretically and empirically, with two other notions of fairness in multi-agent systems: *equality fairness*, where outcomes are distributed so that agents receive as equal outcomes as possible, and *utilitarian fairness*: decisions maximize aggregate welfare, prioritizing total benefit.

Our central claim is that procedural fairness deserves recognition alongside traditional notions of fairness, not as an alternative, but as a principle of legitimacy. We now outline our main contributions:

- We define procedural fairness formally in MA-MABs, and compare it to utilitarian, equality fairness, and Nash welfare.
- We prove impossibility results: fairness notions are fundamentally incompatible, showing that fairness requires normative choices.
- We design algorithms for learning fair policies with sublinear regret guarantees.
- We show that procedurally fair policies lie in the core, ensuring stability against coalitional deviation.
- We empirically evaluate our methods across a variety of settings, and show that procedural fairness balances efficiency and equality while preserving legitimacy.

2 FAIRNESS IN MULTI-AGENT MULTI-ARMED BANDITS

Let N be the number of agents, K be the number of arms, and μ^* be the set of true reward means for each arm and agent, where $\mu^*_{i,k}$ is the true reward mean for agent i when arm k is pulled. We formally define our three notions of fairness, namely, procedural, utilitarian, and equality fairness.

Definition 1 (Procedural Fairness). *Let $P = (p_1, p_2, .., p_k)$ be a policy and let $p_i = \sum_k p_{i,k}$ be the probability mass allocated by agent i across all of the arms. A policy P is procedurally fair if it satisfies the following conditions:*

1. Equal decision-making power. *Each agent $i \in \{1, \ldots, N\}$ is allocated an equal share of the total probability mass, $\sum_{k=1}^{K} p_{i,k} = \frac{1}{N}, \forall i \in \{1, \ldots, N\}$, where $p_{i,k}$ represents the probability mass that agent i contributes to selecting arm k.*

2. Preference-based allocation. *Each agent assigns their probability mass to their most preferred arm(s), defined as the set of arms with the highest mean reward for that agent: $F_i = \{j \in \mathcal{K} \mid \mu^*_{i,j} = \max_{k \in \mathcal{K}} \mu^*_{i,k}\}$. If multiple arms achieve the same maximum expected reward, an agent may distribute their probability mass arbitrarily among them.*

To score a given policy P's procedural fairness, $PF(\mu^*, P)$, we formulate an optimization problem. The intuition is: given some probability distribution, can we allocate $\frac{1}{N}$ of probability on behalf of each agent on their favourite arms, subject to the given policy?

The extent to which we can allocate these decision shares is the procedural fairness score.

Definition 2 (Equality Fairness). *A policy $P = (p_1, \ldots, p_k)$ is equally fair if it minimizes inequality in expected rewards among agents. Formally, the policy P is given by:*

$$P = \arg\min_{p' \in \mathcal{P}} \frac{2}{N(N-1)} \sum_{i>j} \left(\sum_{k=1}^{K} p'_k \mu^*_{i,k} - \sum_{k=1}^{K} p'_k \mu^*_{j,k} \right)^2$$

The equality score, $EF(\mu^*, P)$ for some policy P, is simply the distance of its normalized sum of squared differences from the optimal value.

Definition 3 (Utilitarian Fairness). *A policy $P = (p_1, \ldots, p_k)$ is utilitarian if it maximizes the expected utility among all agents. Formally, the policy P is given by: $P = \arg\max_{p' \in \mathcal{P}} \sum_{i=1}^{N} \sum_{k=1}^{K} p'_k \mu^*_{i,k}$*

The utilitarian score, $UF(\mu^*, P)$ for policy P, is simply the sum of expected utilities across all agents divided by the optimal value.

3 ALGORITHMS

We present learning algorithms for the MA-MAB setting, each optimizing for a specific fairness objective, and also prove regret bounds for each fairness objective. We define regret for procedural fairness as the number of mismatches between estimated and true favourite-arm sets.

To learn a procedurally fair policy, we formulate a constrained optimization problem that ensures each agent allocates an equal decision share to their most preferred arms. When agents have multiple favourite arms, procedural fairness permits many valid allocations. To resolve this ambiguity, we break ties by maximizing decision-share-based Nash welfare, or the product of the sums of the probabilities on each of the agents' favourite arms.

We estimate each agent's favourite-arm set using UCB-style confidence intervals: an arm remains a candidate favourite if its Upper Confidence Bound (UCB) overlaps the Lower Confidence Bound (LCB) of the empirically best arm. To guarantee convergence, we must ensure that these intervals shrink over time, as we need the intervals to converge to 0 to recover the true favourite set with certainty. To solve this problem, we select an arm at random at time t with probability $t^{-(1-\gamma)}$, where $\gamma \in (0, 1)$ is a decay parameter. This guarantees that every arm is pulled sufficiently often so that the confidence radius vanishes as $t \to \infty$.

Theorem 1. *With high probability, the regret bound for the Procedural Fairness algorithm, $R^{PF}(T)$, is $O(T^{\gamma} + [\frac{(1+\alpha)^2 \gamma K \ln(NKT)}{\Delta^2_{min}}]^{\frac{1}{\gamma}})$, where $\Delta_{\min} := \min_{i \in [N]} \min_{j \in F_i} \min_{k \notin F_i} \left(\mu^*_{i,j} - \mu^*_{i,k}\right) > 0$, F_i is the set of agent i's favourite arms based on the true means, and α is an exploration parameter.*

4 THEORETICAL RESULTS

4.1 Procedural Fairness and the Core

The core is a stability notion originating in cooperative game theory [9]. To put it simply in the context of public decision-making [3], it represents a distribution over alternatives (arms) such that no coalition of agents has an incentive to deviate from. In addition to the traditional definition of the core, we introduce a new definition of the core using agents' decision shares (procedural core).

Definition 4 (Procedural Core). *Recall that μ is the reward matrix. Let $F_i = \{k \in \mathcal{K} | \mu^*_{i,k} = \max_{j \in \mathcal{K}} \mu^*_{i,j}\}$ denote agent i's favourite arms. Define a binary vector $X_i \in \{0, 1\}^K$ for each agent i, where $X_i[k]$ is 1 if $k \in F_i$ and 0 otherwise. Thus, given a policy, P, the decision share of agent i is defined as $\beta_i(P) = \sum_{k=1}^{K} X_i[k] p_k$. Same as the outcome core, a policy P is in the core if there is no coalition of agents $A \subseteq \{1, 2, ..., N\}$ and distribution $P' \in \Delta^k$ such that $\frac{|A|}{N} \beta_i(P') \geq \beta_i(P) \quad \forall i \in A$ with at least one strict inequality.*

We now present our main theoretical results:

Theorem 2. *A utility-based Nash Welfare-maximizing distribution [4] need not lie in the procedural core.*

Theorem 3. *With decision-share-based Nash-welfare maximizing tie-breaking, the procedural fairness policy is in the procedural core.*

Theorem 4. *Procedural core implies procedural fairness.*

5 EXPERIMENTS

To understand how our methods work in practice, we conduct experiments on different scenarios and evaluate their performance. We conduct a full factorial sweep across a variety of parameters, resulting in 7,776 different experiment settings. We consider the following notions of fairness: procedural fairness (PF), equality fairness (EF), and utilitarian fairness (UF), as well as the following algorithms: PF, EF, UF policies optimize the fairness notions defined above, NashUCB [4] and an algorithm that optimizes the Generalized Gini Index (GGI) [2]. Table 1 shows the numerical results of the different algorithms in our experiment.

	PF Score	EF Score	UF Score
PF Policy	1.00 ± 0.00	0.98 ± 0.02	0.97 ± 0.05
EF Policy	0.66 ± 0.31	1.00 ± 0.00	0.84 ± 0.13
UF Policy	0.78 ± 0.27	0.96 ± 0.05	1.00 ± 0.00
NSW Policy	0.82 ± 0.23	0.97 ± 0.03	1.00 ± 0.01
GG Policy	0.70 ± 0.28	1.00 ± 0.00	0.87 ± 0.11

Table 1: Performance metrics for each algorithm. Reported as mean ± one standard deviation. Rows denote algorithms' optimal policy, columns denote fairness scores.

6 DISCUSSION & CONCLUSIONS

Fairness in multi-agent learning is typically thought of as a problem of distributing outcomes, like welfare maximization or inequality minimization. However, this overlooks an important dimension of fairness: whether agents have equal influence over the decision-making process itself. In this paper, we argue that procedural fairness deserves recognition alongside other notions of fairness as a principle of legitimacy. Our results formalize procedural fairness in MA-MABs, prove that its policies lie in the procedural core, and show through experiments that it achieves near-optimal performance on common outcome-based fairness objectives. More broadly, this work highlights that fairness objectives are fundamentally incompatible and are, ultimately, normative design choices. Procedural fairness can serve as a principled baseline for settings where perceived legitimacy matters more than any one outcome, and may extend naturally to other decision-making problems.

REFERENCES

[1] Paul Anand. 2001. Procedural fairness in economic and social choice: Evidence from a survey of voters. *Journal of Economic Psychology* 22, 2 (2001), 247–270. https://doi.org/10.1016/S0167-4870(01)00031-9

[2] Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, and Shie Mannor. 2017. Multi-objective bandits: Optimizing the generalized Gini index. In *International Conference on Machine Learning*. PMLR, 625–634.

[3] Brandon Fain, Kamesh Munagala, and Nisarg Shah. 2018. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. 575–592.

[4] Safwan Hossain, Evi Micha, and Nisarg Shah. 2021. Fair algorithms for multi-agent multi-armed bandits. *Advances in Neural Information Processing Systems* 34 (2021), 24005–24017.

[5] Matthew Jones, Huy Nguyen, and Thy Nguyen. 2023. An efficient algorithm for fair multi-agent multi-armed bandit with low regret. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 8159–8167.

[6] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. 2016. Fairness in learning: Classic and contextual bandits. *Advances in Neural Information Processing Systems* 29 (2016).

[7] E.A. Lind and T.R. Tyler. 2013. *The Social Psychology of Procedural Justice*. Springer US. https://books.google.ca/books?id=j97VBQAAQBAJ

[8] John Rawls. 1971. *A Theory of Justice: Original Edition*. Harvard University Press. http://www.jstor.org/stable/j.ctvjf9z6v

[9] Lloyd S Shapley. 1971. Cores of convex games. *International Journal of Game Theory* 1 (1971), 11–26.

[10] T.R. Tyler. 1990. *Why People Obey the Law*. Yale University Press. https://books.google.ca/books?id=ZstoQgAACAAJ