Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Optimality of Independently Randomized Symmetric Policies for Exchangeable Stochastic Teams with Infinitely Many Decision Makers

Sina Sanjari, Naci Saldi, Serdar Yüksel

To cite this article:
Sina Sanjari, Naci Saldi, Serdar Yüksel (2023) Optimality of Independently Randomized Symmetric Policies for Exchangeable Stochastic Teams with Infinitely Many Decision Makers. Mathematics of Operations Research 48(3):1254-1285. https://doi.org/10.1287/moor.2022.1296

https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 48, No. 3, August 2023, pp. 1254–1285
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Optimality of Independently Randomized Symmetric Policies for Exchangeable Stochastic Teams with Infinitely Many Decision Makers

Sina Sanjari,[a,*] **Naci Saldi,**[b] **Serdar Yüksel**[a]

[a] Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada; [b] Department of Mathematics, Bilkent University, 06800 Ankara, Turkey
*Corresponding author

Contact: s.sanjari@queensu.ca, https://orcid.org/0000-0002-6409-7994 (SS); naci.saldi@bilkent.edu.tr, https://orcid.org/0000-0002-2677-7366 (NS); yuksel@queensu.ca, https://orcid.org/0000-0001-6099-5001 (SY)

Received: October 23, 2020
Revised: July 1, 2021; December 31, 2021
Accepted: June 5, 2022
Published Online in Articles in Advance: August 24, 2022

MSC2020 Subject Classification: Primary: 93E20; secondary: 49N80, 60G09

https://doi.org/10.1287/moor.2022.1296

Abstract. We study stochastic teams (known also as decentralized stochastic control problems or identical interest stochastic dynamic games) with large or countably infinite numbers of decision makers and characterize the existence and structural properties of (globally) optimal policies. We consider both static and dynamic nonconvex teams where the cost function and dynamics satisfy an exchangeability condition. To arrive at existence and structural results for optimal policies, we first introduce a topology on control policies, which involves various relaxations given the decentralized information structure. This is then utilized to arrive at a de Finetti–type representation theorem for exchangeable policies. This leads to a representation theorem for policies that admit an infinite exchangeability condition. For a general setup of stochastic team problems with N decision makers, under exchangeability of observations of decision makers and the cost function, we show that, without loss of global optimality, the search for optimal policies can be restricted to those that are N-exchangeable. Then, by extending N-exchangeable policies to infinitely exchangeable ones, establishing a convergence argument for the induced costs, and using the presented de Finetti–type theorem, we establish the existence of an optimal decentralized policy for static and dynamic teams with countably infinite numbers of decision makers, which turns out to be symmetric (i.e., identical) and randomized. In particular, unlike in prior work, convexity of the cost in policies is not assumed. Finally, we show the near optimality of symmetric independently randomized policies for finite N-decision-maker teams and thus establish approximation results for N-decision-maker weakly coupled stochastic teams.

Funding: This work was supported by Natural Sciences and Engineering Research Council of Canada.

Keywords: stochastic teams • mean-field theory • decentralized stochastic control • exchangeable processes

1. Introduction

Stochastic teams consist of a collection of decision makers (DMs) or agents acting together to optimize a common cost function, but not necessarily sharing all the available information. At each time stage, each decision maker only has partial access to the global information, which is defined by the *information structure* (*IS*) of the problem (Witsenhausen [79]). When there is a predefined order according to which the decision makers act, then the team is called a *sequential team*. For sequential teams, if each agent's information depends only on primitive random variables, the team is *static*. If at least one agent's information is affected by an action of another agent, the team is said to be *dynamic*.

In this paper, we study stochastic team problems with large but finite and countably infinite numbers of decision makers. We characterize the existence and structural properties of (globally) optimal policies in such problems. Although teams can be, at first sight, viewed as a narrow class of (identical interest) stochastic dynamic games, when viewed as a generalization of classical single-DM stochastic control, they are quite general, with increasingly common applications involving many areas of applied mathematics, such as decentralized stochastic control (Arrow and Radner [5], Ho [42], Mahajan et al. [60], Sandell et al. [68]), networked control (Hespanha et al. [40], Ho [42]), communication networks (Hespanha et al. [40]), cooperative systems (Beckmann [10], Marschak [61], McGuire [63], Radner [65]), large sensor networks (Tsitsiklis [75]), and energy, or, more specifically, smart grid design (Davison et al. [31], Sandell et al. [68]).

1.1. Connections to Convex Stochastic Teams

For teams with finitely many decision makers, Marschak [61] studied static teams, and Radner [65] established connections between person-by-person optimality, stationarity, and team optimality. Radner's [65] results were generalized in Krainak et al. [52] by relaxing optimality conditions. A summary of these results is that in the context of static teams, the convexity of the cost function, subject to minor regularity conditions, suffices for the global optimality of person-by-person-optimal solutions. In the particular case of linear quadratic Gaussian (LQG) static teams, this result leads to the optimality of linear policies (Radner [65]), which also applies to dynamic LQG problems under partially nested information structures (Ho and Chu [43]). These results are applicable to static teams with finitely many DMs.

In our paper, the main focus is on teams with infinitely many DMs. In this direction, we note that in our prior works (Sanjari and Yüksel [69, 70]), we studied static and dynamic teams where, under convexity and symmetry conditions, global optimality of the limit of the sequence of N-decision-maker optimal policies had been established. These works also provided existence and structural results for convex static and dynamic teams with infinitely many DMs. We also note Mahajan et al. [59], which studied LQG static teams with countably infinite numbers of DMs and established sufficient conditions for global optimality. In our paper here, convexity is not imposed.

1.2. Connections with the Literature on Mean-Field Games/Teams

Team problems can be considered games with identical interests. For the case with infinitely many DMs, a related set of results involves mean-field games: mean-field games (see, e.g., Huang et al. [45, 46], Lasry and Lions [57]) can be viewed as limit models of symmetric non-zero-sum noncooperative finite-player games with a mean-field interaction. We note that in team problems, person-by-person optimality (Nash equilibrium when viewed as games) does not in general imply global optimality for teams with N decision makers and teams with countably infinite numbers of DMs. As we have mentioned, for static teams, a sufficient condition is the convexity of the cost function, subject to minor regularity conditions (Krainak et al. [52]). However, mean-field teams under decentralized information structures generally correspond to dynamic team problems with nonclassical information structures (an observation of a decision maker i is affected by the action of a decision maker j where decision maker i does not have access to the observation of decision maker j); hence, mean-field team problems may be nonconvex even under the convexity of the cost function due to nonclassical information structures (see Yüksel and Saldi [85, section 3.3] and the celebrated counterexample of Witsenhausen [77]). Hence, person-by-person optimality is generally inconclusive for global optimality.

The existence of equilibria for mean-field games was established in Bardi and Fischer [7], Carmona et al. [25], Lasry and Lions [57], Light and Weintraub [58], and Lacker [53]. Furthermore, person-by-person-optimal solutions may perform arbitrarily poorly. There have also been several studies for mean-field games where the limits of sequences of Nash equilibria have been investigated as the number of DMs tends to infinity (see, e.g., Arapostathis et al. [4], Bardi and Priuli [8], Fischer [35], Lacker [56], Lasry and Lions [57]).

Related to mean-field games, in the economic theory literature, Mas-Colell [62] and Schmeidler [73] studied anonymous games and established the existence and characterization of structural properties of Cournot–Nash equilibria. This Cournot–Nash equilibrium concept corresponds to a mean-field equilibrium for a (one-shot) static problem (see also Jovanovic and Rosenthal [48] for sequential anonymous games). However, such an equilibrium does not necessarily imply global optimality in the context of teams because person-by-person optimality (Nash equilibrium when viewed as game problems) does not in general imply global optimality in the absence of convexity. Social optima for mean-field linear quadratic Gaussian control problems under both centralized and restricted decentralized information structures were considered in Arabneydi and Mahajan [3], Huang and Nguyen [44], Huang et al. [47], and Wang and Zhang [76]. We also note the results in Yu et al. [81], where a class of mean-field games, entailing two competitive large teams has been studied. Cecchin [27] studied mean field control problems on a finite state space and showed that the common social optimal expected cost of the N-agent centralized problem (with the classical information structure) converges with an explicit convergence rate to the solution of the corresponding McKean–Vlasov control problem. We refer readers to Caines et al. [19] and Carmona and Delarue [24] for a literature review and a detailed summary of some recent results on mean-field games and social optimum problems.

Some relevant studies on the existence and convergence of equilibria from the mean-field games literature are the following: Cardaliaguet [21], for one-shot mean-field games, under regularity assumptions on the cost function, shows that mixed Nash strategies of N-player symmetric games converge through a subsequence to a limit (which is a weak solution of the mean-field limit). Fischer [35], through a concentration of measures argument, shows that a subsequence of symmetric local approximate Nash equilibria for N-player games converges to a

solution for the mean-field game under the assumption that the normalized occupational measure converges weakly to a deterministic measure. Furthermore, using a similar method, Lacker [54] introduces conditions on equilibrium policies of large-population mean-field symmetric stochastic differential games to allow for convergence of asymmetric approximate Nash equilibria to a weak solution of the mean-field game (Lacker [54, theorem 2.6]) in the presence of common randomness. Using martingale methods and relaxed controls (see also Carmona et al. [25], Fischer [35], Lacker [53, 54]), an existence result and a limit theory are established for controlled McKean–Vlasov dynamics (Lacker [55]). We note that in Lacker [53, 54, 55] and Carmona et al. [25], it is assumed that each player has full access to the information available to all players; that is, the controls are functions of all initial states, Wiener processes of all players, and common randomness.

We further note that the existence results for equilibria have been established in Carmona et al. [25], Carmona and Delarue [24], Fischer [35], and Lacker [54], where strategies of each player are assumed to be progressively measurable to the filtration generated by initial states and Wiener processes (also called *open-loop* controllers in the mean-field games literature (Carmona and Delarue [24], Carmona et al. [25], Fischer [35], Lacker [54]). We note that in our setup, under these strategies, the information structure corresponds to a static information structure. The equilibria with respect to *closed-loop* controllers (in the team problem setup, with respect to a dynamic information structure) can be completely different because the deviating player can still influence the information of the other players, and hence can influence the average of states or actions substantially.

In Lacker [56], under a convexity condition (which was introduced in Filippov [34] and also considered in Lacker [53, 55]) and under the classical information structure (or full information, i.e., what would be a centralized problem in the team theoretic setup), convergence of Nash equilibria induced by (path-dependent and feedback Markovian) closed-loop controllers to a *weak* (*semi-Markov*) *mean-field equilibrium* is established. We also note a result in Cardaliaguet et al. [23] for the convergence of Markov feedback equilibria, where an infinite-dimensional partial differential equation referred to as a *master equation* (obtained as a limit of Hamilton–Jacobi–Bellman equations) is considered and its unique smooth solution used to show the convergence of empirical measures to the unique mean-field game equilibrium. We note that the approach in Cardaliaguet et al. [23] requires uniqueness of the mean-field equilibrium but the one in Lacker [56] applies even if mean-field equilibria are nonunique. In addition, the notion of a weak (semi-Markov) solution considered in Lacker [56] allows for an additional randomization in stochastic flows of measures, but under uniqueness, the limit solution becomes the unique (weak) mean-field equilibrium, and hence recovers the related convergence results in Cardaliaguet et al. [23]. We also note that the convergence problem of Markov feedback equilibria for a finite state model with multiple mean-field equilibria was studied in Hajek and Livesay [38], Bayraktar and Zhang [9], and Cecchin et al. [28]. Recently, in Campi and Fischer [20], both a convergence result for all correlated equilibrium solutions of discrete finite state mean-field games as limits of exchangeable correlated equilibria restricted to Markov open-loop strategies and an approximation result for N-player correlated equilibria were established. For infinite horizon problems, in Cardaliaguet and Rainer [22], an example of ergodic differential games with mean-field coupling is constructed such that limits of sequences of expected costs induced by symmetric Nash equilibria of N-player games capture expected costs induced by many more Nash-equilibrium policies, including a mean-field equilibrium and social optimum. In Lacker [56], the classical information structure (a centralized problem) is considered, whereas in Cardaliaguet and Rainer [22], it is assumed that players have access to the entire history of states of all players but not controls. (We note that in the team setting with the classical information structure, through using a classical result of Blackwell [13] in the case where each DM knows all the history of states of all DMs, optimal policies can be realized as the one in the centralized problem, where just the global state is a sufficient statistic for optimality.) As we see, the information structure aspects can lead to subtle differences in analysis and conclusions.

Furthermore, in the context of stochastic teams with a countably infinite number of DMs, the gap between person-by-person optimality (Nash equilibrium in the game-theoretic context) and global team optimality is significant because a perturbation of finitely many policies fails to deviate the value of the expected cost; thus, person-by-person optimality is a weak condition for such a setup. Hence, without establishing the uniqueness of the mean-field solution (which may hold under strong monotonicity assumptions; Lasry and Lions [57]), the results presented in the aforementioned papers may be inconclusive regarding global optimality of the limit equilibrium. For example, we refer the reader to Bardi and Fischer [7], Cardaliaguet and Rainer [22], and Delarue and Tchuendom [32] for nonuniqueness results and to Hajek and Livesay [38], Bayraktar and Zhang [9] Cecchin et al. [28], and Lacker [56] for connections between nonuniqueness of mean-field equilibria and convergence of Nash equilibria of symmetric N-player games as $N \to \infty$. For team and social optimum control problems, the analysis has primarily focused on the LQG model, where the centralized performance has been shown to be achieved asymptotically by decentralized controllers (see, e.g., Arabneydi and Mahajan [3], Huang et al. [46]).

In this paper, we will adopt a different and novel approach. First, under symmetry of information structures and cost functions, we show that optimal policies are of an exchangeable type for both teams with finite and countably infinite numbers of DMs. Then, in view of our topology on policies, we develop a de Finetti–type representation theorem that characterizes the set of optimal policies as the extreme points of a convex set.

1.3. Connections with Existence Results on Decentralized Stochastic Control

We also note that, compared with the results on the existence of a globally optimal policy for teams with a (finite) number of DMs, N-DM teams (Gupta et al. [37], Saldi [67], Yüksel [83], Yüksel and Saldi [85]), here we study stochastic teams with a countably infinite number of DMs.

In our approach, we use randomized policies for our analysis, and we introduce a topology on control policies for decentralized stochastic control problems. A consequence of our analysis is that, in the limit of countably infinitely many DMs, one can characterize the set of optimal policies as the extreme points of a convex set of policies, which is, in turn, a subset of decentralized, identical, and independently randomized policies (see Theorems 2 and 3). Such a result is not applicable to teams with finitely many DMs. This geometric representation of the set of policies is related to the celebrated de Finetti theorem. De Finetti's theorem implies that infinitely exchangeable joint probability measures can be represented as mixtures (convex combinations) of identical and independent probability measures (Aldous et al. [1], Hewitt and Savage [41], Kingman [51]).

There has been related work in the quantum information/mechanics literature. Let us first note, however, that in Diaconis and Freedman [33], it was shown that a finite number of exchangeable probability measures can be approximated by a mixture of identical and independent probability measures, and this approximation asymptotically becomes more accurate when the number of exchangeable random variables increases. The de Finetti representation–type results have been extended for quantum systems where conditional probability measures have been considered (Banica et al. [6], Brandao and Harrow [17], Caves et al. [26], Christandl and Toner [30], Renner [66]). In fact, for permutation-symmetric conditional probability measures, approximation results have been obtained, provided that the nonsignaling property holds (a conditional independence property between local actions and other measurements given local measurement; Banica et al. [6], Brandao and Harrow [17], Caves et al. [26], Christandl and Toner [30], Renner [66]). We refer readers to Brunner et al. [18] and Popescu [64] for a review of the connection between the nonsignaling conditional probability measures and the conditional probability measures with private and common randomness.

We note that de Finetti–type results developed for conditional probability measures in the quantum information literature give us a geometric interpretation we require for strategic measures (a geometric connection between nonsignaling infinitely exchangeable conditional probability measures and conditional probability measures induced by common and private randomness). However, in the team problem setup, in addition to showing this geometric connection, one is required to show that the common randomness is independent of the observations. We address this issue by introducing an appropriate topology on policies and establishing a de Finetti–type representation theorem on the space of policies, properly defined and metrized.

1.4. Contributions

In view of the above, this paper makes the following contributions:

i. Under symmetry of information structures and exchangeability of the cost function, for teams with N DMs (N-DM teams), we establish the optimality of N-exchangeable randomized policies.

ii. We introduce a suitable topology on control policies that facilitates our analysis using a de Finetti–type representation theorem for decentralized relaxed policies, that is, for the probability measures induced on actions and measurements under decentralized information structures. This leads to a representation theorem for decentralized relaxed policies that admit an infinite exchangeability condition.

iii. By extending N-exchangeable policies to infinitely exchangeable ones, establishing a convergence argument for the induced costs, and using the presented de Finetti theorem for decentralized relaxed policies, we establish the structure and also the existence of optimal decentralized policies for static and dynamic teams with a countably infinite number of DMs, which turns out to be symmetric (i.e., identical) and randomized. Compared with our previous results for static and dynamic mean-field teams in Sanjari and Yüksel [70, theorem 12 or proposition 1] and Sanjari and Yüksel [69, theorem 3.4], (i) the cost function is not necessarily convex in actions, (ii) action spaces are not necessarily convex, and (iii) the mean-field coupling is considered in dynamics, which leads to a nonclassical information structure (a consequence being that the problem is in general nonconvex in policies).

iv. For N–decision maker symmetric teams with a symmetric information structure, we show that symmetric (identical) randomized policies of mean-field teams are nearly optimal.

2. Preliminaries and Statement of Main Results

We begin with Witsenhausen's intrinsic model for team problems and then provide a description of the main problems studied in this paper.

2.1. Preliminaries

In this section, we introduce Witsenhausen's intrinsic model for sequential teams (Witsenhausen [79]):

• There exists a collection of *measurable spaces* $\{(\Omega,\mathcal{F}),(\mathbb{U}^i,\mathcal{U}^i),(\mathbb{Y}^i,\mathcal{Y}^i),i\in\mathcal{N}\}$, specifying the system's distinguishable events and control and measurement spaces. The set $\mathcal{N}$ denotes the collection of DMs. The set $\mathcal{N}$ can be a finite set $\{1,2,\dots,N\}$ or a countable set $\mathbb{N}$. The pair $(\Omega,\mathcal{F})$ is a measurable space (on which an underlying probability may be defined). The pair $(\mathbb{U}^i,\mathcal{U}^i)$ denotes the standard Borel space from which the action u^i of DM^i is selected. The pair $(\mathbb{Y}^i,\mathcal{Y}^i)$ denotes the standard Borel observation/measurement space for each decision maker i (DM^i).

• There is a *measurement constraint* to establish the connection between the observation variables and the system's distinguishable events. The $\mathbb{Y}^i$-valued observation variables are given by $y^i=h^i(\omega,\underline{u}^{[1,i-1]})$, where $\underline{u}^{[1,i-1]}:=(u^1,\dots,u^{i-1})$, and h^i's are measurable functions.

• The set of admissible control laws $\underline{\gamma}:=(\gamma^i)_{i\in\mathcal{N}}$, also called *designs* or *policies*, are measurable control functions, so that $u^i=\gamma^i(y^i)$. Let Γ^i denote the set of all admissible policies for DM^i, and let $\Gamma=\prod_{i\in\mathcal{N}}\Gamma^i$. These policies will later be allowed to be randomized, and, accordingly, the image will be $\mathcal{P}(\mathbb{U}^i)$, where $\mathcal{P}(\cdot)$ denotes the space of probability measures.

• There is a *probability measure* $\mathbb{P}$ on $(\Omega,\mathcal{F})$ describing the probability space on which the system is defined.

Under this intrinsic model, a sequential team problem is *dynamic* if the information available to at least one DM is affected by the action of at least one other DM. A team problem is *static* if for every DM the information available is affected only by exogenous disturbances, that is, no other DM can affect the information at any given DM. Information structures can also be categorized as *classical*, *quasi-classical*, or *nonclassical*. An IS $\{y^i,i\in\mathcal{N}\}$ is classical if y^i contains all of the information available to DM^k for $k<i$. An IS is quasi-classical or *partially nested* if, whenever u^k, for some $k<i$, affects y^i through the measurement function h^i, y^i contains y^k (i.e., $\sigma(y^k)\subset\sigma(y^i)$). An IS that is not partially nested is *nonclassical*.

In the paper, we will also allow for randomized policies, where in addition to y^i, each DM^i has access to common and private randomization. This will be made precise later in Section 3.1.

2.2. Problem Statement

We consider stochastic team problems with finite but large and team problems with countably infinite numbers of DMs. We address three main problems: the (i) existence and structural results for static teams with a countably infinite number of DMs (Section 4), (ii) the existence and structural results for dynamic teams with a countably infinite number of DMs (Section 5), and (iii) approximation results for N-DM static and dynamic teams (Section 6).

2.2.1. Static Teams.

As we consider exchangeable team problems, we let action and observation spaces be identical across DMs $\mathbb{U}^i=\mathbb{U}\subseteq\mathbb{R}^n$ and $\mathbb{Y}^i=\mathbb{Y}\subseteq\mathbb{R}^m$ for all $i\in\mathcal{N}$, where n and m are positive integers.

Problem ($\mathcal{P}_N$). Let $\mathcal{N}=\{1,\dots,N\}$. Let $\underline{\gamma}_N:=(\gamma^1,\dots,\gamma^N)$ and $\Gamma_N:=\prod_{i=1}^N\Gamma^i$. Let an expected cost function of $\underline{\gamma}_N$ be given by

$$J_N(\underline{\gamma}_N)=\mathbb{E}^{\underline{\gamma}_N}[c(\omega_0,\underline{u}_N)]:=\mathbb{E}\big[c(\omega_0,\gamma^1(y^1),\dots,\gamma^N(y^N))\big],\tag{1}$$

for some Borel measurable cost function $c:\Omega_0\times\prod_{k=1}^N\mathbb{U}\to\mathbb{R}_+$. We define ω_0 as the Ω_0-valued, cost-function-relevant, exogenous random variable as $\omega_0:(\Omega,\mathcal{F},\mathbb{P})\to(\Omega_0,\mathcal{F}_0)$, where Ω_0 is a Borel space with its Borel σ-field $\mathcal{F}_0$. Here, we have the notation $\underline{u}_N:=(u^1,\dots,u^N)$.

Definition 1. For a given stochastic team problem $(\mathcal{P}_N)$ with a given information structure, a policy (strategy) $\underline{\gamma}_N^*:=(\gamma^{1*},\dots,\gamma^{N*})\in\Gamma_N$ is (*globally*) *optimal* for $(\mathcal{P}_N)$ if

$$J_N(\underline{\gamma}_N^*)=\inf_{\underline{\gamma}_N\in\Gamma_N}J_N(\underline{\gamma}_N).$$

Our focus in this paper is on a class of exchangeable team problems satisfying an exchangeability assumption on the cost function.

Assumption 1. *The cost function is exchangeable with respect to actions for all ω_0, that is, for any permutation σ of $\{1,\dots,N\}$, $c(\omega_0,u^1,\dots,u^N)=c(\omega_0,u^{\sigma(1)},\dots,u^{\sigma(N)})$ for all ω_0.*

In particular, for our main results, we focus on team problems with the following expected cost function instead of (1):

$$\mathbb{E}^{\underline{\gamma}_N}\left[\frac{1}{N}\sum_{i=1}^{N}c\left(\omega_0,u^i,\frac{1}{N}\sum_{p=1}^{N}u^p\right)\right]. \tag{2}$$

Clearly, $\frac{1}{N}\sum_{i=1}^{N}c(\omega_0,u^i,\frac{1}{N}\sum_{p=1}^{N}u^p)$ satisfies Assumption 1, and the cost function c in (2), in fact, can be made more general by considering any cost function $c:\Omega_0\times\mathbb{U}\times\tilde{\mathbb{U}}\to\mathbb{R}_+$, depending on the empirical measure of the control actions under mild continuity conditions, that is,

$$c\left(\omega_0,u^i,\Xi\left(\frac{1}{N}\sum_{p=1}^{N}\delta_{u^p}\right)\right), \tag{3}$$

where $\Xi:\mathcal{P}(\mathbb{U})\to\tilde{\mathbb{U}}$ is weakly continuous, $\delta_{(\cdot)}$ is a Dirac delta measure, and $\tilde{\mathbb{U}}$ is a subset of appropriate dimensional Euclidean space. However, for clarity in presentation, we will follow (2).

Next, we introduce a stochastic team problem with a countably infinite number of DMs.

Problem ($\mathcal{P}_\infty$). Consider a stochastic team with a countably infinite number of DMs, that is, $\mathcal{N}=\mathbb{N}$. Let $\Gamma:=\prod_{i\in\mathbb{N}}\Gamma^i$ and $\underline{\gamma}:=(\gamma^1,\gamma^2,\dots)$. Let an expected cost of $\underline{\gamma}$ be given by

$$J(\underline{\gamma})=\limsup_{N\to\infty}\mathbb{E}^{\underline{\gamma}}\left[\frac{1}{N}\sum_{i=1}^{N}c\left(\omega_0,u^i,\frac{1}{N}\sum_{p=1}^{N}u^p\right)\right], \tag{4}$$

for some Borel measurable cost function $c:\Omega_0\times\mathbb{U}\times\mathbb{U}\to\mathbb{R}_+$.

Definition 2. For a given stochastic team problem $(\mathcal{P}_\infty)$ with a given information structure, a policy $\underline{\gamma}^*:=(\gamma^{1*},\gamma^{2*},\dots)\in\Gamma$ is (*globally*) *optimal* for $(\mathcal{P}_\infty)$ if

$$J(\underline{\gamma}^*)=\inf_{\underline{\gamma}\in\Gamma}J(\underline{\gamma}).$$

Later on, we allow DMs to apply randomized policies and provide a description of the problems within randomized policies; see (20) and (21). Our first goal here is to establish the existence of a symmetric (identical) randomized globally optimal policy for static mean-field team problems $(\mathcal{P}_\infty)$. To this end, we first establish N-exchangeability of randomized optimal policies for $(\mathcal{P}_N)$ and symmetry for optimal randomized policies of $(\mathcal{P}_\infty)$. Then, in Theorem 2, using symmetry, we establish an existence result for $(\mathcal{P}_\infty)$. Our theorems require the following absolute continuity condition under which we can equivalently view the observations of each DM as independent and also independent of ω_0 via change of measure argument (due to Witsenhausen [78]).

Assumption 2. *Assume that for every $N\in\mathbb{N}\cup\{\infty\}$, there exists a probability measure Q^i on $\mathbb{Y}$ and a function f^i for all $i\in\mathcal{N}$ such that for all Borel set B^i in $\mathbb{Y}$ (with $B:=B^1\times\cdots\times B^N$),*

$$\tilde{\mu}^N(B|\omega_0)=\prod_{i=1}^{N}\int_{B^i}f^i(y^i,\omega_0,y^1,\dots,y^{i-1})Q^i(dy^i), \tag{5}$$

where $\tilde{\mu}^N$ is the conditional distribution of observations $(y^1,\dots,y^N)$ given ω_0.

Remark 1. In particular, if y^i takes values from a countable set, Assumption 2 always holds, for example, with the reference measure taken as $Q^i(r)=\sum_{p\geq1}2^{-p}1_{\{r=m_p\}}$, where $\mathbb{Y}=\{m_p|p\in\mathbb{N}\}$, and $1_{\{\cdot\}}$ denotes the indicator function (see Witsenhausen [78]).

The above allows us to introduce a suitable topology under which the space of randomized policies is Borel (see Section 3.1). In addition, our main Theorem 2 imposes the following assumptions on the observations and action spaces.

Assumption 3. *Assume the following:*
i. $(y^i)_{i\in\mathcal{N}}$ *are independent and identically distributed (i.i.d.), conditioned on* ω_0;
ii. $\mathbb{U}$ *is compact.*

We note that under Assumptions 2 and 3i, there exists an identical reference probability measure Q and function f such that the absolute continuity condition (5) holds so that for any Borel set B^i in $\mathbb{Y}$ (with $B := B^1 \times \cdots \times B^N$),

$$\begin{aligned}\tilde{\mu}^N(B|\omega_0) &= \prod_{i=1}^N \hat{\mu}(B^i|\omega_0)\\ &= \prod_{i=1}^N \int_{B^i} f(y^i,\omega_0)Q(dy^i),\end{aligned}$$

where $\hat{\mu}$ is the conditional distribution of each observation y^i given ω_0. We note that the function f and the measure Q are identical across DMs because observations are identically distributed conditioned on ω_0. This change of measure argument allows us to equivalently rewrite the expected cost function with respect to the underlying probability measure $\mathbb{P}$ as a new cost function $\tilde{c}$ with respect to a probability measure $\mathbb{P}$, where observations are independent (also independent of ω_0; see Witsenhausen [78] and Yüksel [83, section 2.2]):

$$E^{\gamma^N}_{\mathbb{P}}[c(\omega_0,u^1,\dots,u^N)] = E^{\gamma^N}_{\mathbb{Q}}[\tilde{c}(\omega_0,y^1,\dots,y^N,u^1,\dots,u^N)], \tag{6}$$

where $\mathbb{Q}(A\times\prod_{i=1}^N B^i) := \int_{A\times\prod_{i=1}^N B^i}\prod_{i=1}^N Q(dy^i)\mathbb{P}^0(d\omega_0)$ for any Borel set $D = A\times\prod_{i=1}^N B^i$ in $\Omega_0\times\prod_{i=1}^N\mathbb{Y}$, $\mathbb{P}^0$ is the distribution of ω_0, and $\tilde{c}$ is given by

$$\tilde{c}(\omega_0,y^1,\dots,y^N,u^1,\dots,u^N) := c(\omega_0,u^1,\dots,u^N)\prod_{i=1}^N f(y^i,\omega_0).$$

Furthermore, our main Theorem 2 imposes the following continuity assumption on the cost function.

Assumption 4. *The cost function in* (2), $c:\Omega_0\times\mathbb{U}\times\mathbb{U}\to\mathbb{R}_+$, *is continuous in its second and third arguments for all* ω_0.

For our results in Section 4, we impose Assumptions 1 and 2, but we impose Assumptions 3 and 4 only when they are needed.

2.2.2. Dynamic Teams. Our second goal here is to establish the existence of a symmetric (identical) randomized globally optimal policy for mean-field dynamic team problems where DMs are weakly coupled through the average of states and actions in dynamics and/or the cost function. Again, we consider exchangeable teams, and, hence, we let action, observation, and state spaces, respectively, be identical across DMs $i\in\mathcal{N}$ and, for simplicity, also through time; $t = 0,\dots,T-1$, $\mathbb{U}^i_t = \mathbb{U}\subseteq\mathbb{R}^n$, $\mathbb{Y}^i_t = \mathbb{Y}\subseteq\mathbb{R}^{n'}$, $\mathbb{X}^i_t = \mathbb{X}\subseteq\mathbb{R}^{n''}$ for all $i\in\mathcal{N}$ and $t = 0,\dots,T-1$, where n, n', and n'' are positive integers. Define state dynamics and observation dynamics of DMs as follows:

$$x^i_{t+1} = f_t\left(x^i_t,u^i_t,\frac{1}{N}\sum_{p=1}^N x^p_t,\frac{1}{N}\sum_{p=1}^N u^p_t,w^i_t\right), \tag{7}$$

$$y^i_t = h_t(x^i_{0:t},u^i_{0:t-1},v^i_{0:t}), \tag{8}$$

where functions f_t and h_t are measurable functions, and v^i_t and w^i_t are random vectors representing uncertainties in state dynamics and observations. We have the notations $x^i_{0:t} := (x^i_0,\dots,x^i_t)$, $u^i_{0:t-1} := (u^i_0,\dots,u^i_{t-1})$, and $v^i_{0:t} := (v^i_0,\dots,v^i_t)$. Let the admissible policies $(\gamma^i_{0:T-1})_{i\in\mathcal{N}}$ (with $\gamma^i_{0:T-1} := (\gamma^i_0,\dots,\gamma^i_{T-1})$) be measurable control functions so that $u^i_t = \gamma^i_t(y^i_t)$ for all $i\in\mathcal{N}$ and $t = 0,\dots,T-1$. We note that the state dynamics (7) can be made more general by considering any measurable function f_t (for $t = 0,\dots,T-1$), depending on the empirical measures of the states and control actions under mild continuity conditions, that is,

$$f_t\left(x^i_t,u^i_t,\Xi^x\left(\frac{1}{N}\sum_{p=1}^N\delta_{x^p_t}\right),\Xi^u\left(\frac{1}{N}\sum_{p=1}^N\delta_{u^p_t}\right),w^i_t\right), \tag{9}$$

where $\Xi^u : \mathcal{P}(\mathbb{U}) \to \tilde{\mathbb{U}}$ and $\Xi^x : \mathcal{P}(\mathbb{X}) \to \tilde{\mathbb{X}}$ are weakly continuous, and $\tilde{\mathbb{U}}$ and $\tilde{\mathbb{X}}$ are subsets of appropriate dimensional Euclidean spaces.

Problem ($\mathcal{P}_T^N$). Consider N-DM mean-field dynamic teams with the expected cost function of $\underline{\gamma}^{1:N}$ as

$$J_T^N(\underline{\gamma}^{1:N}) = \mathbb{E}^{\underline{\gamma}^{1:N}}\left[\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N} c\left(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\right)\right], \tag{10}$$

where $\underline{\gamma}^{1:N} := (\gamma^1_{0:T-1}, \dots, \gamma^N_{0:T-1})$ and $\gamma^i_{0:T-1} := (\gamma^i_0, \dots, \gamma^i_{T-1})$. Again, $\omega_0 : (\Omega, \mathcal{F}, \mathbb{P}) \to (\Omega_0, \mathcal{F}_0)$ is a cost-related random variable, where Ω_0 is a Borel space with its Borel σ-field $\mathcal{F}_0$.

Problem ($\mathcal{P}_T^\infty$). Consider mean-field dynamic teams with the expected cost function of $\underline{\gamma}$ as

$$J_T^\infty(\underline{\gamma}) = \limsup_{N\to\infty} J_T^N(\underline{\gamma}^{1:N}), \tag{11}$$

where $\underline{\gamma} := (\gamma^1_{0:T-1}, \gamma^2_{0:T-1}, \dots)$ and $\underline{\gamma}^{1:N} := (\gamma^1_{0:T-1}, \dots, \gamma^N_{0:T-1})$.

Again, the cost function c in (10) can be made more general by considering any cost function c as follows:

$$c\left(\omega_0, x_t^i, u_t^i, \Xi^u\left(\frac{1}{N}\sum_{p=1}^{N}\delta_{u_t^p}\right), \Xi^x\left(\frac{1}{N}\sum_{p=1}^{N}\delta_{x_t^p}\right)\right), \tag{12}$$

where Ξ^u and Ξ^x are weakly continuous. Again, for clarity in presentation, we will follow the setting with the state dynamics (7) and the cost function c in (10). Analogous to Definition 1 and Definition 2, we can define globally optimal policies for ($\mathcal{P}_T^N$) and ($\mathcal{P}_T^\infty$). Again, we allow DMs to apply randomized policies and provide a description of the problems within randomized policies; see (25) and (26). In Section 5, we establish the existence of a symmetric (identical across DMs) randomized globally optimal policy for ($\mathcal{P}_T^\infty$). Similar to the static case, we first establish N-exchangeablity of randomized optimal policies for ($\mathcal{P}_T^N$) and symmetry for randomized optimal policies of ($\mathcal{P}_T^\infty$). Then, using symmetry, we establish an existence result for ($\mathcal{P}_T^\infty$).

Our solution technique for dynamic problems is similar to that for static ones, which requires more technical arguments and additional assumptions. Our theorems for the dynamic case impose an absolute continuity condition (see Assumption 8) that allows us to equip control policies with a suitable topology and facilitates our analysis (our main Theorem 3 requires an additional technical assumption, Assumption 11). Furthermore, our main Theorem 3 imposes the following.

Assumption 5. *Assume the following:*

i. *For $t = 0, \dots, T-1$, functions f_t and h_t in (7) and (8) are continuous in the states and actions, and f_t's are bounded.*

ii. *The cost function in (10), $c : \Omega_0 \times \mathbb{X} \times \mathbb{U} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}_+$, is continuous in the second, third, fourth, and fifth arguments.*

Assumption 6. *Assume the following:*

i. *$(x_0^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors conditioned on ω_0.*

ii. *For $t = 0, \dots, T-1$, $(w_t^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors, and for $i \in \mathcal{N}$, $(w_t^i)_{t=0}^{T-1}$ are mutually independent, and independent of ω_0 and $(x_0^i)_{i\in\mathcal{N}}$. For $t = 0, \dots, T-1$, $(v_t^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors, and for $i \in \mathcal{N}$, $(v_t^i)_{t=0}^{T-1}$ are mutually independent, and independent of ω_0, $(x_0^i)_{i\in\mathcal{N}}$, and w_t^i's for $i \in \mathcal{N}$ and $t = 0, \dots, T-1$.*

iii. *$\mathbb{U}$ is compact.*

In view of Assumption 6i, we note that ω_0 also introduces a correlation between initial states. For our results in Section 5, we impose Assumption 8, and we impose Assumptions 11, 5, and 6 only when they are needed.

2.2.3. Approximations. Finally, we establish the following approximations in Section 6. If P_π^* is a (randomized) symmetric optimal policy for ($\mathcal{P}_\infty$) (($\mathcal{P}_T^\infty$)), then there exist $\epsilon_N \geq 0$, with $\epsilon_N \to 0$ as $N \to \infty$, such that $P_\pi^*|_N$ is ϵ_N-optimal for ($\mathcal{P}_N$) (($\mathcal{P}_T^N$)), where $P_\pi^*|_N$ is the restriction of P_π^* to the first N DMs. For this, we use our symmetry results and analysis for ($\mathcal{P}_\infty$) (($\mathcal{P}_T^\infty$)).

2.3. Discussion of the Main Results

In mean-field team problems, one may be interested in the existence and structure of globally optimal policies. In particular, one can ask whether there is a globally optimal policy, and whether this optimal policy is symmetric

for this class of problems (by a symmetric policy, we mean that a policy is identical across DMs). One may also be interested in the connection between optimal policies for mean-field teams and approximations for optimal policies of the prelimit N-DM teams, when N is large. The goal of this paper is to address these questions for mean-field team problems, where the problem can be nonconvex. The nonconvexity of the problem can arise as a result of nonconvexity of the action space and/or nonconvexity of the cost function in actions. Also, even if the action space is convex and the cost function is convex in actions, the information structure of the problem may lead to nonconvexity of the problem in policies (see, e.g., Yüksel and Saldi [85, section 3.3]). A celebrated example is the counterexample of Witsenhausen [77].

One of the main difficulties in studying nonconvex mean-field teams is to show that a symmetric (identical for each DM) globally optimal policy exists. This difficulty stems from the observation that, in general, globally optimal policies are not symmetric for nonconvex prelimit N-DM team problems (which can be seen in Example 1). This is in contrast to convex mean-field teams, where symmetry can be established for both prelimit N-DM and mean-field team problems (Sanjari and Yüksel [69, 70]). Our approach is as follows:

i. We introduce a topology on control policies that is used to establish a de Finetti representation result for probability measures on policies identified as randomized policies. In Theorem 1, we show that any infinitely exchangeable randomized policies can be represented by elements of the set of randomized policies with common and private independent randomness, where, conditioned on common randomness, randomization of the policies is independent and identical across DMs.

ii. In Section 4 for static and Section 5 for dynamic N-DM stochastic teams (see Lemma 1 and Lemma 3), we show that by exchangeability of the cost function and considering symmetric information structures (under a causality condition for the dynamic case), one can establish N-exchangeability of randomized optimal policies.

iii. In Section 4 for static and Section 5 for dynamic mean-field teams (see Lemma 2 and Lemma 4) under regularity conditions on the cost function and dynamics, by constructing infinitely exchangeable randomized policies by relabeling N-exchangeable randomized optimal policies, as N goes to infinity, we show the asymptotic optimality of infinitely exchangeable randomized optimal policies. Hence, this, following from our de Finetti–type theorem (see Theorem 1), establishes asymptotic global optimality of symmetric and conditionally independent policies.

iv. Using extreme point and lower semicontinuity arguments, we establish the existence of a symmetric optimal policy (which is privately randomized) for static and dynamic mean-field teams (see Theorem 2 and Theorem 3).

v. In Section 6, using our analysis for mean-field problems, as N goes to infinity, we show that symmetric optimal policies of mean-field teams are asymptotically optimal for N-DM weakly coupled teams; hence, it establishes approximation results for this class of problems.

In the following, we first study static teams, and then we study dynamic teams, where the analysis is similar to that for the static case but is somewhat more technical.

3. Topology on Control Policies and a de Finetti Representation Result

3.1. Topology on Control Policies

In this section, we introduce a topology, using which we can introduce Borel probability measures on policies. We first consider N-DM static team problems. Following from Witsenhausen [78] and Yüksel [83], Assumption 2 allows us to reduce the problem as a static team problem where now the observation of each DM is independent of observations of other DMs and also independent of ω_0 (because under the measure transformation (5), a probability measure on the observation of each DM is Q^i, which is independent of observations of other DMs and ω_0). Hence, under Assumption 2, we can focus on each DM^i separately. Let us define

$$\Theta^i := \left\{ P \in \mathcal{P}(\mathbb{U} \times \mathbb{Y}) \,\middle|\, P(B) = \int_B 1_{\{g^i(y^i) \in du^i\}} Q^i(dy^i),\ g^i : \mathbb{Y} \to \mathbb{U},\ B \in \mathcal{B}(\mathbb{U} \times \mathbb{Y}) \right\}, \tag{13}$$

where $\mathcal{P}(\cdot)$ denotes the space of probability measures, and $1_{\{\cdot \in A\}}$ denotes the indicator function of the set A. The above set is the set of extreme points of the set of probability measures on $\mathbb{U} \times \mathbb{Y}$ with fixed marginals Q^i on $\mathbb{Y}$, that is,

$$\mathcal{R}^i := \left\{ P \in \mathcal{P}(\mathbb{U} \times \mathbb{Y}) \,\middle|\, P(B) = \int_B \Pi^i(du^i | y^i) Q^i(dy^i),\ B \in \mathcal{B}(\mathbb{U} \times \mathbb{Y}) \right\}, \tag{14}$$

where Π^i is a stochastic kernel from $\mathbb{Y}$ to $\mathbb{U}$. Hence, Θ^i inherits Borel measurability and topological properties of the Borel measurable set $\mathcal{R}^i$ (Borkar [16]). We note that this set corresponds to Young [80] measures, and this

representation result is due to Borkar [16]. Now, we identify the set of relaxed policies Γ^i by $\mathcal{R}^i$, and we define convergence on policies as $\gamma^i_n \to \gamma^i$ if and only if $\gamma^i_n(du^i|y^i)Q^i(dy^i) \to \gamma^i(du^i|y^i)Q^i(dy^i)$ (in the weak convergence topology) as $n \to \infty$.

In view of the above standard Borel space formulation for Γ^i for each $i \in \mathcal{N}$, we can define the set of Borel probability measures on admissible policies $\Gamma_N := \Gamma^1 \times \cdots \times \Gamma^N$ (which we will refer to as a set of randomized policies) as $L^N := \mathcal{P}(\Gamma_N)$, where Borel σ-field $\mathcal{B}(\Gamma^i)$ is induced by the topology defined above. Define the set of randomized policies induced by common and individual randomness as

$$L^N_{\mathrm{CO}} := \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \int_{z\in[0,1]} \prod_{i=1}^N P^i_\pi(\gamma^i \in A_i|z)\eta(dz), \quad \eta \in \mathcal{P}([0,1]) \Bigg\},$$

where η is the distribution of common but independent (from intrinsic exogenous system variables) randomness. In the above, for every fixed z, $P^i_\pi \in \mathcal{P}(\Gamma^i)$ indicates an independent randomized policy for each DM^i ($i = 1, \dots, N$). Note that, conditioned on a $[0,1]$-valued random variable Z, policies are independent. It can be shown that L^N_{CO} and L^N are identical (see Theorem A.1 in Appendix A), and hence, the set of randomized policies L^N corresponds to randomized policies induced by individual and common randomness. Because individual and common randomness do not improve the optimal expected cost, the relaxation of the problem to sets of randomized policies L^N is a legitimate relaxation for the team problems with N DMs.

Before we introduce the set of exchangeable randomized policies, we recall the definition of *exchangeability* for random vectors.

Definition 3. Random vectors $x^1, x^2, \dots, x^N$ defined on a common probability space are N-exchangeable if for any permutation σ of the set $\{1, \dots, N\}$,

$$\mathcal{L}\left(x^{\sigma(1)}, x^{\sigma(2)}, \dots, x^{\sigma(N)}\right) = \mathcal{L}\left(x^1, x^2, \dots, x^N\right),$$

where $\mathcal{L}$ denotes the joint distribution of random vectors. Random vectors $(x^1, x^2, \dots)$ are infinitely exchangeable if finite distributions of $(x^1, x^2, \dots)$ and $(x^{\sigma(1)}, x^{\sigma(2)}, \dots)$ are identical for any finite permutation (affecting only finitely many elements) of $\mathbb{N}$.

Now, we define the set of exchangeable randomized policies as

$$L^N_{\mathrm{EX}} := \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) \text{ and for all } \sigma \in S_N : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = P_\pi(\gamma^{\sigma(1)} \in A_1, \dots, \gamma^{\sigma(N)} \in A_N) \Bigg\}, \tag{15}$$

where S_N is the set of all permutations of $\{1, \dots, N\}$. We note that L^N_{EX} is a convex subset of L^N. We also define the set $L^N_{\mathrm{CO,SYM}}$ as the set of identical randomized policies induced by a common and individual randomness:

$$L^N_{\mathrm{CO,SYM}} := \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \int_{z\in[0,1]} \prod_{i=1}^N \tilde{P}_\pi(\gamma^i \in A_i|z)\eta(dz), \quad \eta \in \mathcal{P}([0,1]) \Bigg\},$$

where for all $i \in \mathcal{N}$ and fixed z, $\tilde{P}_\pi \in \mathcal{P}(\Gamma^i)$ indicates an identical independent randomized policy for each DM^i ($i = 1, \dots, N$). Also, define the set of randomized policies with only private independent randomness as

$$L^N_{\mathrm{PR}} := \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \prod_{i=1}^N P^i_\pi(\gamma^i \in A_i), \text{ for } P^i_\pi \in \mathcal{P}(\Gamma^i) \Bigg\}.$$

Finally, define the set of randomized policies with identical and independent randomness as

$$L^N_{\mathrm{PR,SYM}} := \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \prod_{i=1}^N \tilde{P}_\pi(\gamma^i \in A_i), \quad \text{for } \tilde{P}_\pi \in \mathcal{P}(\Gamma^i) \Bigg\}. \tag{16}$$

For a team with a countably infinite number of DMs, we define sets of randomized policies $L, L_{\mathrm{EX}}, L_{\mathrm{CO}}, L_{\mathrm{CO,SYM}}$, $L_{\mathrm{PR}}, L_{\mathrm{PR,SYM}}$ similarly using the Ionescu Tulcea extension theorem through the sequential formulation reviewed in Section 2.1, by iteratively adding new coordinates for our probability measure (see, e.g., Aliprantis and Border [2], Hernández-Lerma and Lasserre [39]). We define the set of randomized policies L on the infinite product Borel spaces $\Gamma := \prod_{i\in\mathbb{N}} \Gamma^i$ as $L := \mathcal{P}(\Gamma)$. Now, let the set of all infinitely exchangeable randomized policies be given by

$$L_{\mathrm{EX}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) \text{ and for all } N \in \mathbb{N}, \text{and for all } \sigma \in S_N : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = P_\pi(\gamma^{\sigma(1)} \in A_1, \dots, \gamma^{\sigma(N)} \in A_N) \Bigg\},$$

and let the set of all randomized policies with common and independent randomness be given by

$$L_{\mathrm{CO}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} P^i_\pi(\gamma^i \in A_i | z) \eta(dz), \quad \eta \in \mathcal{P}([0,1]) \Bigg\}.$$

Note that L_{CO} is a convex subset of L, and its extreme points are in the set of randomized policies with private independent randomness:

$$L_{\mathrm{PR}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \prod_{i\in\mathbb{N}} P^i_\pi(\gamma^i \in A_i), \quad \text{for } P^i_\pi \in \mathcal{P}(\Gamma^i) \Bigg\}.$$

Also, we define

$$L_{\mathrm{CO,SYM}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i | z) \eta(dz), \quad \eta \in \mathcal{P}([0,1]) \Bigg\},$$

and

$$L_{\mathrm{PR,SYM}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i), \quad \text{for } \tilde{P}_\pi \in \mathcal{P}(\Gamma^i) \Bigg\}.$$

3.2. A de Finetti Theorem for Admissible Team Policies

In view of the introduced topology and sets of Borel probability measures on policies (sets of randomized policies), we now establish a connection between L_{EX} and $L_{\mathrm{CO,SYM}}$ using the classical de Finetti theorem; that is, infinitely exchangeable randomized policies are a mixture of i.i.d. randomized policies.

Theorem 1. *Any infinitely exchangeable randomized policy $P_\pi \in L_{\mathrm{EX}}$ is in the set of randomized policies $L_{\mathrm{CO,SYM}}$ ($P_\pi \in L_{\mathrm{CO,SYM}}$); that is, for any $P_\pi \in L_{\mathrm{EX}}$, there exists a $[0,1]$-valued random variable Z such that for any $A_i \in \mathcal{B}(\Gamma^i)$,*

$$P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i | z)\eta(dz), \quad \eta \in \mathcal{P}([0,1]), \tag{17}$$

where for every fixed z, $\tilde{P}_\pi \in \mathcal{P}(\Gamma^i)$.

Proof. In view of the introduced weak convergence topology on Γ^i (using Borel measurable sets (14) and (13)), we have that Γ^i is a closed subset of the Borel space $\mathcal{P}(\mathbb{U}\times\mathbb{Y})$, and, hence, Γ^i is Borel for $i \in \mathbb{N}$. The proof follows from Kallenberg [50, theorem 1.1] because $\Gamma = \prod_{i=1}^\infty \Gamma^i$ is Borel. We note that the de Finetti representation in Kallenberg [50, theorem 1.1] is of the form $P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{\mathcal{P}(\Gamma^i)} \prod_{i=1}^\infty m(A^i)\hat{\eta}(dm)$ for $\hat{\eta} \in \mathcal{P}(\mathcal{P}(\Gamma^i))$, which can be written as in (17). That is because $\mathcal{P}(\Gamma^i)$ is an (uncountable) Borel space (Bertsekas and Shreve [12, corollary 7.25.1]), and, hence, by the Borel-isomorphism theorem (see, e.g., Bertsekas and Shreve [12, proposition 7.16]), it is Borel isomorphic to $[0,1]$. □

4. Existence and Structure of Optimal Policies for Symmetric Static Team Problems with Infinitely Many DMs

In this section, we consider static stochastic team problems where we impose Assumptions 1 and 2. All the proofs for this section are presented in Appendix B. Based on the definitions of randomized policies, we redefine the expected cost in $(\mathcal{P}_N)$ of a randomized policy $P_\pi \in L^N$ as

$$\begin{aligned} J_N^\pi(\underline{\gamma}_N) &:= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &:= \int\left(\int c(\omega_0, u^1, \dots, u^N)\prod_{k=1}^N \gamma^k(du^k|y^k)\right)P_\pi(d\gamma^1, \dots, d\gamma^N)\mu^N(d\omega_0, dy^1, \dots, dy^N), \end{aligned} \tag{18}$$

where $c^N(\underline{\gamma}, \underline{y}, \omega_0) := \int c(\omega_0, u^1, \dots, u^N)\prod_{k=1}^N \gamma^k(du^k|y^k)$, and μ^N is the joint probability measure on measurements $(y^1, \dots, y^N)$ and ω_0. In the following, we characterize team problems in which the search for a randomized optimal policy can be restricted to policies in L^N_{EX} without losing global optimality.

Assumption 7. *Assume $(y^1, \dots, y^N)$ are exchangeable, conditioned on ω_0.*

Note that Assumption 7 is weaker than Assumption 3i.

Lemma 1. *For a fixed N, consider an N-DM static team. Assume $\bar{L}^N$ is an arbitrary convex subset of L^N. If Assumption 7 holds, then*

$$\inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) = \inf_{P_\pi \in \bar{L}^N \cap L^N_{\mathrm{EX}}} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0). \tag{19}$$

In the following, we present an existence result on globally optimal policies for static mean-field teams with infinitely many DMs. First, we restate the mean-field team problem and its prelimit under randomized policies.

Problem $(\mathcal{P}_N)$. Consider an N-DM static team with the following expected cost of a randomized policy $P^N_\pi \in L^N$:

$$\begin{aligned} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) := \int\left(\int \frac{1}{N}\sum_{i=1}^N c\left(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p\right)\prod_{k=1}^N \gamma^k(du^k|y^k)\right) \\ \times P^N_\pi(d\gamma^1, \dots, d\gamma^N)\mu^N(d\omega_0, dy^1, \dots, dy^N). \end{aligned} \tag{20}$$

The above problem is considered as a prelimit problem for our infinite-DM team problem. This problem is a special case of $(\mathcal{P}_N)$ defined in the previous section because we have a special structure for the cost function c^N which satisfies Assumption 1.

Problem $(\mathcal{P}_\infty)$**.** Consider a infinite-DM static team with the following expected cost of a randomized policy $P_\pi \in L$:

$$\limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0), \tag{21}$$

where $P_{\pi,N}$ is the marginal of the $P_\pi \in L$ to the first N components, and μ^N is the marginal of the fixed probability measure on $(\omega_0, y^1, y^2, \dots)$ to the first $N+1$ components.

In the following, we present a key result required for our main theorem. Under mild conditions, we show that the optimal expected costs induced by L_{EX}^N and L_{EX} are equal as N goes to infinity. Hence, by Lemma 1, under symmetry, this allows us to show that, without loss of global optimality, optimal policies of static mean-field teams with a countably infinite number of DMs can be considered to be of an infinitely exchangeable type.

Lemma 2. *Suppose that Assumptions 3 and 4 hold. Assume further that the cost function is bounded. Then*

$$\begin{aligned}&\limsup_{N\to\infty} \inf_{P_\pi^N \in L_{\mathrm{EX}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\ &\quad= \limsup_{N\to\infty} \inf_{P_\pi \in L_{\mathrm{EX}}} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0),\end{aligned} \tag{22}$$

where $P_{\pi,N}$ is the marginal of the $P_\pi \in L_{\mathrm{EX}}$ to the first N components.

In the following, we establish the existence of a randomized optimal policy for $(\mathcal{P}_\infty)$.

Theorem 2. *Consider a static team problem $(\mathcal{P}_\infty)$ where Assumptions 3 and 4 hold. Then, there exists a randomized optimal policy P_π^* for $(\mathcal{P}_\infty)$ that is in $L_{\mathrm{PR,SYM}}$:*

$$\begin{aligned}&\inf_{P_\pi \in L_{\mathrm{PR,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\ &\quad= \limsup_{N\to\infty} \int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\ &\quad= \inf_{P_\pi \in L_{\mathrm{PR}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0).\end{aligned}$$

Remark 2. We note that Lemma 1, Lemma 2, and Theorem 2 apply even if the more general cost function (3) is considered. Lemma 1 applies using an argument identical to that used in the proof of Lemma 1 because the cost function (3) satisfies Assumption 1, and Lemma 2 applies by an argument identical to that used in the proof of Lemma 2 (by (weak) continuity of Ξ, we can establish a counterpart of (B.10), and hence, Step 2 in the proof of Lemma 2 follows identically). Theorem 2 applies by an argument identical to that used in the proof of Theorem 2, using (weak) continuity of Ξ.

In the following, we present an example where Theorem 2 can be applied but the existence result of Sanjari and Yüksel [70, theorem 12] cannot be applied because the assumption that $\mathbb{U}$ is convex in Sanjari and Yüksel [70, theorem 12] is violated.

Example 1. Consider a team problem with the following expected cost function:

$$J(\underline{\gamma}) = \limsup_{N\to\infty} \mathbb{E}^{\underline{\gamma}}\left[\left(\left(\frac{1}{N}\sum_{i=1}^N u^i\right) - \frac{1}{2}\right)^2\right],$$

where σ-field $\sigma(y^i) = \{\emptyset, \Omega\}$ (this corresponds to a team setup where DMs have no measurement, hence measurable functions (policies) are constant functions). Let $\mathbb{U}^i = \{0,1\}$ for each DM. Clearly, an optimal policy that achieves zero is the one where there is a matching partition (such as even numbers versus odd numbers) among DMs picking $u^i = 0$ and $u^i = 1$, because the cost function is nonnegative. One can see that there is an optimal policy in $L_{\mathrm{PR,SYM}}$ because each DM can choose independently an action zero or one with probability one-half, and this achieves the expected cost of zero; however, there is no identically deterministic policy that achieves zero expected cost. We note also that since the action sets are not convex, the results in Sanjari and Yüksel [70, theorem 12 or proposition 1] are not applicable and hence can not guarantee the existence of a symmetric randomized optimal policy, in particular, the action sets are not convex.

5. Finite Horizon Dynamic Team Problems with a Symmetric Information Structure

In this section, we study dynamic stochastic teams. All the proofs for this section are presented in Appendix C. As for the static case, we first introduce the intrinsic model for general dynamic teams, and then we introduce a topology on control policies. Finally, we establish our main results for dynamic problems.

5.1. A Revised Intrinsic Model for Dynamic Team Problems

Under the intrinsic model (see Section 2.1), every DM acts separately. However, depending on the information structure, it may be convenient to consider a collection of DMs as a single DM acting at different time instances. In fact, in the classical stochastic control problem, this is the standard approach. In this subsection, we introduce the general (multistage) dynamic teams using the intrinsic model under deterministic policies. In the next subsections, we allow randomization equipped with a suitable topology.

According to the discussion above, by considering a collection of DMs as a single DM ($i = 1, \dots, N$) acting at different time instances ($t = 0, \dots, T-1$), we revise the intrinsic mode for (multistage) dynamic teams with (NT) DMs as a team with N DMs (for $N \in \mathbb{N} \cup \{\infty\}$):

i. Let the observation and action spaces be standard Borel spaces and be identical for each DM ($i = 1, \dots, N$) with $\mathbb{Y}_i := \mathbf{Y} = \prod_{t=0}^{T-1} \mathbb{Y}^t$, $\mathbb{U}_i := \mathbf{U} = \prod_{t=0}^{T-1} \mathbb{U}^t$, respectively. (Later on, for simplicity of our notation and analysis, we assume that action and observation spaces are also identical through time.) For each DM^i, the set of all admissible policies is denoted by $\Gamma_i := \prod_{t=0}^{T-1} \Gamma^t$. Later on, these policies will be allowed to be randomized and, accordingly, the image will be $\mathcal{P}(\mathbf{U})$.

ii. For $i = 1, \dots, N$, $y_t^i := h_t^i(x_0^{1:N}, \zeta_{0:t}^{1:N}, u_{0:t-1}^{1:N})$ represents the observation of DM^i at time t (h_t^i's are Borel measurable functions). Let ν_t^N be a stochastic kernel characterizing the joint distribution of observations $y_t^{1:N} := (y_t^1, \dots, y_t^N)$ at time t induced by h_t^i's given the available information, and let $(\underline{\zeta}^{1:N}) := (\underline{\zeta}^1, \dots, \underline{\zeta}^N)$, where $\underline{\zeta}^i := (x_0^i, \zeta_{0:T-1}^i)$ denotes all the uncertainty associated with DM^i, including his or her initial states. We assume that each $\underline{\zeta}^i$ takes a value in $\Omega_{\underline{\zeta}}$ (where at each time instance t, it takes a value in Ω_{ζ_t}). Let μ^N denote the joint distribution of random variables $\underline{\zeta}^{1:N}$. To be consistent with our notations in our analysis of the static case, we use the same notation μ^N for the fixed probability measures on observations, and ω_0 for the static case. However, we note that in the dynamic case, the probability measures on uncertainties $\underline{\zeta}^{1:N}$ are fixed, but probability measures on observations are not fixed.

5.2. Topology on Dynamic Control Policies

Similar to Section 3.1, here, we allow randomization in policies, but first we introduce two reduction conditions (independent and nested reduction) that enable us to define sets of Borel probability measures on randomized policies for dynamic teams with different information structures.

Assumption 8. *One of the following conditions holds:*

i. *Independent reduction: For every $N \in \mathbb{N} \cup \{\infty\}$ and for $i = 1, \dots, N$ and $t = 0, \dots, T-1$, there exists a probability measure τ_t^i on $\mathbb{Y}^t$ and a function $\psi_t^i : \mathbb{Y}^t \times \Omega_0 \times \prod_{p=1}^N \left(\prod_{k=0}^{t-1} \Omega_{\zeta_k} \times \prod_{k=0}^{t-1} (\mathbb{U}^k \times \mathbb{Y}^k) \right) \to \mathbb{R}_+$ such that for all Borel sets A^i on $\mathbb{Y}^t$ (with $A = A^1 \times \cdots \times A^N$),*

$$\nu_t^N\left(A \,|\, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right) = \prod_{i=1}^N \int_{A^i} \psi_t^i\left(y_t^i, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right) \tau_t^i(dy_t^i).$$

ii. *Nested reduction: For every $N \in \mathbb{N} \cup \{\infty\}$ and for $i = 1, \dots, N$ and $t = 0, \dots, T-1$, there exists a probability measure η_t^i on $\mathbb{Y}^t$ and a function ϕ_t^i such that for all Borel sets A^i on $\mathbb{Y}^t$ (with $A = A^1 \times \cdots \times A^N$),*

$$\begin{aligned} &\nu_t^N\left(A \,|\, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right) \\ &= \prod_{i=1}^N \int_{A^i} \phi_t^i\left(y_t^i, \omega_0, x_0^{-i}, \zeta_{0:t-1}^{-i}, y_{0:t-1}^{-i}, u_{0:t-1}^{-i}\right) \eta_t^i\left(dy_t^i \,|\, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i\right), \end{aligned}$$

and for each DM^i through time ($t = 0, \dots, T-1$), there exists a static reduction with the classical information structure (i.e., under the reduction, the information structure of each DM through time is expanding such that $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$ for $t = 0, \dots, T-1$).

We note that Assumption 8i allows us to obtain independent measurement reductions both across DMs and through time, $t = 0, \dots, T-1$ (see Appendix C, Section C.1). Assumption 8ii holds if an independent static reduction exists across DMs and there exists a nested static reduction for each DM through time; that is, under the reduction, the information is expanding for each DM through time (see Appendix C, Section C.1). In view of the above reduction conditions, we introduce a suitable topology for randomized policies. Similar to Section 3.1, under Assumption 8i, we define convergence on policies as

$$\underline{\gamma}^i_n \overset{n\to\infty}{\longrightarrow} \underline{\gamma}^i \quad \text{if and only if} \quad \gamma^i_{t,n}(du^i_t|y^i_t)\tau^i_t(dy^i_t) \underset{\text{weakly}}{\overset{n\to\infty}{\longrightarrow}} \gamma^i_t(du^i_t|y^i_t)\tau^i_t(dy^i_t) \quad \forall t = 0, \dots, T-1.$$

Under Assumption 8ii, we define convergence on policies as

$$\underline{\gamma}^i_n \overset{n\to\infty}{\longrightarrow} \underline{\gamma}^i \quad \text{if and only if} \quad \gamma^i_{t,n}(du^i_t|y^i_{0:t})\eta^i_t(dy^i_{0:t}) \underset{\text{weakly}}{\overset{n\to\infty}{\longrightarrow}} \gamma^i_t(du^i_t|y^i_{0:t})\eta^i_t(dy^i_{0:t}) \quad \forall t = 0, \dots, T-1.$$

Hence, under Assumption 8, we can define all the sets of randomized policies introduced in Section 3.1 for dynamic teams by considering $\underline{\gamma}^i$.

Remark 3. We note that our first reduction condition, independent reduction, is essentially a version of Girsanov's [36] transformation (Beneš [11]), which was considered first in Witsenhausen [78, equation (4.2)], and later utilized in Yüksel S, Başar [84, p. 114] and Yüksel [83, section 2.2]. (For discrete-time partially observed stochastic control, similar arguments have been presented, for example, by Borkar [14, 15].) We refer the reader to Charalambous [29] for relations with the classical continuous-time stochastic control, where the relation with Girsanov's [36] classical measure transformation (Beneš [11]) is recognized. Our second reduction condition, nested reduction, holds when there exists a reduction for DMs through time under which each DM has perfect recall of a private history of information.

Now, we provide examples under which either one of the conditions in Assumption 8 holds.

Example 2. For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x^i_{t+1} = f^i_t(x^{1:N}_{0:t}, u^{1:N}_{0:t}, w^i_t)$ and $y^i_t = h^i_t(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}) + v^i_t$, where $\zeta^i_t := (w^i_t, v^i_t)$, and v^i_t admits zero-mean Gaussian density function θ^i_t with positive-definite covariance. Then,

i. if the information structure for each DM at time t is described as $I^i_t := \{y^i_t\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 8i holds;

ii. if $I^i_t := \{y^i_{0:t}, u^i_{0:t-1}\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$ (or, equivalently, $I^i_t := \{\tilde{y}^i_t\}$ with $\tilde{y}^i_t := \tilde{h}^i_t(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}, v^i_{0:t})$ for some function $\tilde{h}^i_t$ and $\sigma(\tilde{y}^i_t) \subset \sigma(\tilde{y}^i_{t+1})$ and $\sigma(u^i_t) \subset \sigma(\tilde{y}^i_{t+1})$ for some function $\bar{h}_t$), then Assumption 8ii holds.

Part i is true because for all $t = 0, \dots, T-1$ and $i = 1, \dots, N$, we have

$$y^i_t = h^i_t\Big(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\Big) + v^i_t = \kappa^i_t\Big(\omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\Big) + v^i_t,$$

for some functions κ^i_t, and hence, we can define

$$\psi^i_t\Big(y^i_t, \omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, y^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\Big) := \frac{\theta^i_t\Big(y^i_t - \kappa^i_t\Big(\omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\Big)\Big)}{\theta^i_t(y^i_t)}, \ \tau^i_t(dy^i_t) := \theta^i_t(y^i_t)dy^i_t,$$

where dy^i_t is with respect to the Lebesgue measure. Part ii can be shown similarly by first applying the independent reduction as above among DMs, and then considering the nested information structure through time for each DM.

Example 3. Consider the following two information structures:

i. (Open-loop information structure) For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x^i_{t+1} = f^i_t(x^{1:N}_{0:t}, u^{1:N}_{0:t}, w^i_t)$ and $y^i_t = h^i_t(\zeta^i_{0:t-1}, v^i_t)$ such that $\sigma(y^i_t) \subset \sigma(y^i_{t+1})$, where $(\zeta^i_t)_t := (w^i_t, v^i_t)_t$ denotes the disturbances of DM^i (which are independent of disturbances of other DMs and independent of ω_0). If $I^i_t := \{y^i_t\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 8ii holds.

ii. For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x^i_{t+1} = f^i_t(\omega_0, x^{1:N}_{0:t}, u^{1:N}_{0:t}) + w^i_t$, where w^i_t admits zero-mean Gaussian density function θ^i_t with positive-definite covariance, and let $y^i_t = h^i_t(x^i_{0:t}, y^i_{0:t-1}, v^i_{0:t})$ such that $\sigma(y^i_t) \subset \sigma(y^i_{t+1})$, where

$(v_t^i)_t$ are independent of disturbances of other DMs and independent of ω_0. If $I_t^i := \{y_t^i\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 8ii holds.

Part i follows from the fact that the information structure is open-loop and nested for each DM, and, hence, under this information structure, the problem is static with the classical information structure through time for each DM. Part ii is true because for all $t = 0, \dots, T-1$ and $i = 1, \dots, N$,

$$\hat{\phi}_t^i\left(x_t^i, \omega_0, x_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right) := \frac{\theta_t^i(x_t^i - f_t^i(\omega_0, x_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}))}{\theta_t^i(x_t^i)}, \ \hat{\eta}_t^i(dx_t^i) := \theta_t^i(x_t^i)dx_t^i,$$

and because the information structure is nested through time for each DM.

5.3. Existence and Structure of Optimal Policies for Symmetric Dynamic Teams with Infinitely Many DMs

In the following, we study the existence and structure of globally optimal policies for dynamic teams with a symmetric information structure (that are not necessarily partially nested) and with a finite but large and also infinitely many DMs. We note that a related result is given in Sanjari and Yüksel [69], where convex mean-field teams are studied under the assumption that the action space is convex for each DM and the cost function is convex in policies. We note that even if the cost function is convex in actions when there is a mean-field coupling in dynamics, convexity rarely holds because the information structure under a decentralized setup is nonclassical, and that may lead to the nonconvexity of the team problem in policies (see, e.g., Yüksel and Saldi [85, section 3.3]). In the following, convexity is not imposed. Again, for our results in this subsection, we impose Assumption 8.

5.3.1. Exchangeability of Optimal Policies for Symmetric Dynamic Teams with a Finite but Large Number of DMs.

In this subsection, we focus on symmetric dynamic teams with N DMs, and we establish a structural result for optimal policies of this class of problems (which is more general than the prelimit mean-field model ($\mathcal{P}_T^N$)). In the next subsection, we use this result to establish the existence and structural properties of globally randomized optimal policies for mean-field dynamic team problems.

Now, we recall the definition of the symmetric information structure from Sanjari and Yüksel [69] (note that symmetric information structures can be classical, partially nested, or nonclassical). Several examples as well as a graph interpretation of dynamic teams with symmetric information structures are presented in Sanjari and Yüksel [69, section 4]. In particular, prelimit mean-field and mean-field dynamic teams ($\mathcal{P}_T^N$) and ($\mathcal{P}_T^\infty$), introduced in Section 2.2, have a symmetric information structure.

Definition 4 (Sanjari and Yüksel [69]). Let the information of DM^i acting at time t be described as $I_t^i := \{y_t^i\}$. The information structure of a sequential N-DM team problem is *symmetric* if $y_t^i = h_t(x_0^i, x_0^{-i}, \zeta_{0:t}^i, \zeta_{0:t}^{-i}, u_{0:t-1}^i, u_{0:t-1}^{-i})$, where h_t is identical for all $i = 1, \dots, N$ (note that the arguments of the function depend on i) and $b^{-i} = (b^1, \dots, b^{i-1}, b^{i+1}, \dots, b^N)$ for $b = x_0, \zeta_{0:t}, u_{0:t-1}$.

We note that the above definition can be generalized to be applicable for teams with a countably infinite number of DMs. Before we present the result for dynamic mean-field teams, we characterize team problems with symmetric information structures in which the search for an optimal policy can be restricted to policies in L_{EX}^N without losing global optimality. To this end, we focus on a more general setup of team problems within randomized policies $P_\pi \in L^N$ as

$$\begin{aligned} J_N^\pi(\underline{\gamma}^{1:N}) &:= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\ &:= \int\left(\int c(\omega_0, \underline{\zeta}^{1:N}, \underline{u}^1, \dots, \underline{u}^N)\prod_{i=1}^N \underline{\gamma}^i(d\underline{u}^i|\underline{y}^i)\right)P_\pi(d\underline{\gamma}^1, \dots, d\underline{\gamma}^N)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\ &\quad \times \prod_{t=0}^{T-1} v_t^N\left(dy_t^{1:N}|\omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right), \end{aligned} \tag{23}$$

where $c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) := \int c(\omega_0, \underline{\zeta}^{1:N}, \underline{u}^1, \dots, \underline{u}^N)\prod_{i=1}^N \underline{\gamma}^i(d\underline{u}^i|\underline{y}^i)$ and the following assumptions hold.

Assumption 9. *For any permutation σ of the set $\{1,\dots,N\}$, we have, for all ω_0,*

$$c\Big(\omega_0,(\underline{\zeta}^{\sigma})^{1:N},(\underline{u}^{\sigma})^{1:N}\Big)=c\Big(\omega_0,\underline{\zeta}^{1:N},\underline{u}^{1:N}\Big), \tag{24}$$

where $(\underline{\zeta}^{\sigma})^{1:N}:=(\underline{\zeta}^{\sigma(1)},\dots,\underline{\zeta}^{\sigma(N)})$ and $(\underline{u}^{\sigma})^{1:N}:=(\underline{u}^{\sigma(1)},\dots,\underline{u}^{\sigma(N)})$.

Assumption 10. *Assume the following:*

i. *$(\underline{\zeta}^1,\dots,\underline{\zeta}^N)$ are exchangeable conditioned on ω_0;*

ii. *for all $t=0,\dots,T-1$, and all Borel sets A^i on $\mathbb{Y}^t$ (with $A=A^1\times\cdots\times A^N$),*

$$\nu_t^N\Big(A\Big|\omega_0,x_0^{1:N},\zeta_{0:t-1}^{1:N},y_{0:t-1}^{1:N},u_{0:t-1}^{1:N}\Big)=\prod_{i=1}^{N}\nu_t^i\Big(A^i\Big|\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^{1:N},u_{0:t-1}^{1:N}\Big),$$

where ν_t^i is a stochastic kernel of the observation DM^i at time t, y_t^i, induced by h_t (which is identical for each DM).

We note that dynamic mean-field team problems introduced in Section 2.2.2, with the cost function as (10), state dynamics as (7), and observations as (8), under Assumption 6 satisfy Assumptions 9 and 10.

Lemma 3. *Consider a dynamic team with a symmetric information structure. Let Assumptions 9 and 10 hold. Suppose further that $\bar{L}^N$ is an arbitrary nonempty convex subset of L^N. Then*

$$\inf_{P_\pi\in\bar{L}^N}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$=\inf_{P_\pi\in\bar{L}^N\cap L^N_{\mathrm{EX}}}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0).$$

5.3.2. Existence and Structure of Optimal Policies for Mean-Field Dynamic Teams. In the following, we establish the existence of a globally randomized optimal policy for dynamic mean-field teams with infinitely many DMs. Define state dynamics and observations as (7) and (8), respectively. The information structure of DM^i at time t is $I_t^i=\{y_t^i\}$, and $\zeta_t^i:=(w_t^i,v_t^i)$ (with $\zeta_0^i:=(x_0^i,w_0^i,v_0^i)$) denotes the uncertainty corresponding to dynamics and observations at time t for DM^i, which are exogenous random vectors in standard Borel spaces. First, we reformulate the mean-field team problem and its prelimit within randomized policies.

Problem ($\mathcal{P}_T^N$). Consider an N-DM dynamic team with the expected cost of a randomized policy $P_\pi^N\in L^N$ as

$$\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$:=\int\Bigg(\int\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N}c\Bigg(\omega_0,x_t^i,u_t^i,\frac{1}{N}\sum_{p=1}^{N}u_t^p(y_t^p),\frac{1}{N}\sum_{p=1}^{N}x_t^p\Bigg)\prod_{k=1}^{N}\underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)\Bigg)$$

$$\times P_\pi^N(d\underline{\gamma}^1,\dots,d\underline{\gamma}^N)\mu^N(d\omega_0,d\underline{\zeta}^{1:N})\prod_{t=0}^{T-1}\nu_t^N\Big(dy_t^{1:N}|\omega_0,x_0^{1:N},\zeta_{0:t-1}^{1:N},y_{0:t-1}^{1:N},u_{0:t-1}^{1:N}\Big), \tag{25}$$

where

$$c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0):=\int\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N}c\Bigg(\omega_0,x_t^i,u_t^i,\frac{1}{N}\sum_{p=1}^{N}u_t^p(y_t^p),\frac{1}{N}\sum_{p=1}^{N}x_t^p\Bigg)\prod_{k=1}^{N}\underline{\gamma}^k(d\underline{u}^k|\underline{y}^k).$$

The above problem is considered as a prelimit problem for the mean-field team problem. We note that $(\mathcal{P}_T^N)$ is a special case of (23) because we have a special structure for the cost function c^N and observations, which satisfy Assumption 9 and Definition 4, respectively.

Remark 4. Our analysis below also allows more general observations for each DM, where the observations of each DM at time t can be explicitly functions of averages of previous states and actions as

$$y_t^i = h_t\left(x_{0:t}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^N x_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, v_{0:t}^i\right).$$

However, to simplify the presentations of theorems and proofs and emphasize the decentralization of the optimal policy, for this rest of the paper, we consider (8).

Problem $(\mathcal{P}_T^\infty)$**.** Consider an infinite-DM static team with the following expected cost of a randomized policy $P_\pi \in L$:

$$\limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0), \tag{26}$$

where $P_{\pi,N}$ is the restriction of $P_\pi \in L$ to its first N components, and μ^N is the marginal of the fixed probability measure on $(\omega_0, \underline{\zeta}^1, \underline{\zeta}^2, \dots)$ to the first $N+1$ components.

Assumption 11. *Assumption 8 holds with functions ψ_t^i and ϕ_t^i of the following forms for every $i \in \mathcal{N}$ and $t = 0, \dots, T-1$:*

$$\psi_t^i\left(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N x_{0:t}^p\right),$$

$$\phi_t^i\left(y_t^i, \omega_0, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N x_{0:t}^p\right),$$

where ψ_t^i is continuous in the last three arguments (actions and the empirical means of actions and states), and ϕ_t^i is continuous in the last two arguments (the empirical means of actions and states).

Before presenting our main result for dynamic mean-field teams, we introduce sufficient conditions under which the expected cost function induced by randomized optimal policies in L_{EX}^N and L_{EX} are equal as N goes to infinity, and hence, following from Lemma 3, under symmetry, this shows that without loss of global optimality, optimal policies of dynamic mean-field teams can be considered to be of an infinitely exchangeable type.

Lemma 4. *Consider the team problem $(\mathcal{P}_T^N)$, where Assumption 11, Assumption 5, and Assumption 6 hold. Assume further that the cost function is bounded. Then*

$$\limsup_{N\to\infty} \inf_{P_\pi^N \in L_{\mathrm{EX}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$= \limsup_{N\to\infty} \inf_{P_\pi \in L_{\mathrm{EX}}} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0), \tag{27}$$

where $P_{\pi,N}$ is the restriction of $P_\pi \in L_{\mathrm{EX}}$ to its first N components, and μ^N is the marginal of the fixed probability measure on $(\omega_0, \underline{\zeta}^1, \underline{\zeta}^2, \dots)$ to the first $N+1$ components.

In the following, we establish an existence and structural result for a randomized optimal policy of $(\mathcal{P}_T^\infty)$.

Theorem 3. *Consider a mean-field team $(\mathcal{P}_T^\infty)$ with $(\mathcal{P}_T^N)$ having a symmetric information structure for every N. Let Assumption 11, Assumption 5, and Assumption 6 hold. Then, there exists a randomized optimal policy P_π^* for (P_T^∞) that is in $L_{\mathrm{PR,SYM}}$,*

$$\inf_{P_\pi \in L_{\mathrm{PR,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$:= \limsup_{N\to\infty} \int P_{\pi,N}^*(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$= \inf_{P_\pi \in L} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0).$$

Remark 5. Similar to Remark 2 for the static teams, we can show that for dynamic teams, Lemma 3, Lemma 4, and Theorem 3 apply for a more general setting described by the state dynamics as (9) and the cost function as (12).

6. Approximations of Optimal Policies for Symmetric *N*-DM Stochastic Teams

In this section, we present approximation results for optimal policies of N-DM teams. We show that for large N, symmetric policies are nearly optimal, and the restriction of the optimal infinite solution is nearly optimal for the finite team when N is large. Proofs of the theorems in this section are provided in Appendix D. We first consider the static case. To present ideas more effectively, we first introduce the following set of probability measures on policies:

$$L_{\mathrm{D}}^{N} := \left\{ P_\pi \in L^N \;\middle|\; \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \ldots, \gamma^N \in A_N) = \prod_{i=1}^{N} 1_{\{\tilde{\gamma}^i \in A_i\}}, \;\; \text{for } \tilde{\gamma}^i \in \Gamma^i \right\},$$

where the above set corresponds to Dirac delta measures in L_{PR}^{N}.

Theorem 4. *Consider a static team $(\mathcal{P}_N)$ (see (20)), where Assumption 4 and Assumption 3 hold. Suppose further that the cost function is bounded. Then,*

i.

$$\inf_{P_\pi^N \in L_{\mathrm{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi^N \in L_{\mathrm{CO}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N, \tag{28}$$

and

$$\inf_{P_\pi^N \in L_{\mathrm{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi^N \in L_{\mathrm{D}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N, \tag{29}$$

where $\epsilon_N \to 0$ as N goes to infinity;

ii. *if $P_\pi^* \in L_{\mathrm{PR,SYM}}$ is a randomized optimal policy of $(\mathcal{P}_\infty)$, then there exist $\bar{\epsilon}_N \geq 0$ where $\bar{\epsilon}_N \to 0$ as N goes to infinity and*

$$\int P_{\pi,N}^*(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi^N \in L_{\mathrm{D}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N + \bar{\epsilon}_N, \tag{30}$$

where $P_{\pi,N}^*$ is the restriction of P_π^* to the first N components.

The main idea for establishing part i is to use Lemmas 1 and 2 to provide an approximation of optimal expected cost by restricting the search for randomized policies to those that are restrictions of randomized policies in L_{EX} to the N first components. We note that because the set of policies L_{D}^N is not a convex subset of the set of randomized policies L^N, (28) does not immediately imply (29) using Lemma 1; however, the result can be established using an extreme point argument and the fact that policies in L_{D}^N are optimal among all randomized policies L_{PR}^N for N-DM teams (thanks to Blackwell's [13] irrelevant information theorem). Part ii follows from part i and Theorem 2, using the fact that a randomized optimal policy $P_\pi^* \in L_{\mathrm{PR,SYM}}$ of $(\mathcal{P}_\infty)$ provides an approximation for the optimal expected cost when the search for randomized optimal policy for N-DM teams is restricted to those in $L_{\mathrm{PR,SYM}}^N$.

Similarly, we present approximation results for optimal policies of symmetric dynamic N-DM teams.

Theorem 5. *Consider a dynamic team $(\mathcal{P}_T^N)$ (see (10)). Let Assumption 11, Assumption 5, and Assumption 6 hold. If the cost function is bounded, then*

i.

$$\inf_{P_\pi^N \in L_{\mathrm{PR,\,SYM}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{\zeta}) c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) \nu^N(d\underline{y} | \underline{\zeta}, \underline{\gamma}, \omega_0)$$

$$\leq \inf_{P_\pi^N \in L_{\mathrm{CO}}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{\zeta}) c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) \nu^N(d\underline{y} | \underline{\zeta}, \underline{\gamma}, \omega_0) + \epsilon_N, \tag{31}$$

and

$$\inf_{P_\pi^N \in L^N_{\mathrm{PR,\,SYM}}} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$
$$\leq \inf_{P_\pi^N \in L^N_{\mathrm{D}}} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0) + \epsilon_N, \tag{32}$$

where $\epsilon_N \to 0$ as N goes to infinity;

ii. *if $P_\pi^* \in L_{\mathrm{PR,SYM}}$ is a randomized optimal policy for $(\mathcal{P}_T^\infty)$, then there exist $\bar{\epsilon}_N \geq 0$ where $\bar{\epsilon}_N \to 0$ as N goes to infinity and*

$$\int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$
$$\leq \inf_{P_\pi^N \in L^N_{\mathrm{D}}} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0) + \epsilon_N + \bar{\epsilon}_N,$$

where $P^*_{\pi,N}$ is the restriction of P^*_π to the first N components.

Proof. The proof follows from steps similar to those used in the proof of Theorem 4 by using the results of Lemma 4 and Theorem 3. □

Appendix A. Connection Between L^N_{CO} and L^N in Section 3.1

In following theorem, we show that sets of randomized policies L^N_{CO} and L^N are identical.

Theorem A.1. *The set of randomized policies L^N is identical to the set of randomized policies L^N_{CO}.*

Proof. Clearly, we have $L^N_{\mathrm{CO}} \subseteq L^N$. In the following, we show that $L^N \subseteq L^N_{\mathrm{CO}}$. Following from Borkar [16], for each $i = 1, \ldots, N$, the set of marginals on each coordinate Γ^i of randomized policies belonging to L^N is a convex combination of its extreme points which is a subset of the set of Dirac delta measures of elements in Γ^i. Hence, we have

$$\mathrm{Extreme}(L^N) \subseteq \left\{ P_\pi \in L^N \;\middle|\; \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \ldots, \gamma^N \in A_N) = \prod_{i=1}^N 1_{\{\tilde{\gamma}^i \in A_i\}}, \;\; \text{for } \tilde{\gamma}^i \in \Gamma^i \right\},$$

where $\mathrm{Extreme}(L^N)$ denotes the set of extreme points of the convex set L^N. Hence, $\mathrm{Extreme}(L^N) \subseteq L^N_{\mathrm{CO}}$. Because L^N and L^N_{CO} are convex, we have that $L^N \subseteq L^N_{\mathrm{CO}}$, and this completes the proof. □

Appendix B. Proofs from Section 4

B.1. Proof of Lemma 1

For any permutation $\sigma \in S_N$, we define a randomized policy $P^\sigma_\pi \in \bar{L}^N$ as a permutation, σ, of arguments of a randomized policy $P_\pi \in \bar{L}^N$; that is, for $A^i \in \mathcal{B}(\Gamma^i)$,

$$P^\sigma_\pi(\gamma^1 \in A^1, \ldots, \gamma^2 \in A^N) := P_\pi(\gamma^{\sigma(1)} \in A^1, \ldots, \gamma^{\sigma(N)} \in A^N). \tag{B.1}$$

We have

$$\int P^\sigma_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)$$
$$= \int c(\omega_0, u^1, \ldots, u^N)\prod_{k=1}^N \gamma^k(du^k|y^k)\tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0)P^\sigma_\pi(d\gamma^1, \ldots, d\gamma^N)\mathbb{P}_0(d\omega_0)$$
$$= \int c(\omega_0, u^1, \ldots, u^N))\prod_{k=1}^N \gamma^k(du^k|y^k)\tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0)P_\pi(d\gamma^{\sigma(1)}, \ldots, d\gamma^{\sigma(N)})\mathbb{P}_0(d\omega_0) \tag{B.2}$$
$$= \int c(\omega_0, u^{\sigma(1)}, \ldots, u^{\sigma(N)}))\prod_{k=1}^N \gamma^{\sigma(k)}(du^{\sigma(k)}|y^{\sigma(k)})\tilde{\mu}^N(dy^{\sigma(1)}, \ldots, dy^{\sigma(N)}|\omega_0)P_\pi(d\gamma^1, \ldots, d\gamma^N)\mathbb{P}_0(d\omega_0) \tag{B.3}$$
$$= \int c(\omega_0, u^1, \ldots, u^N)\prod_{k=1}^N \gamma^k(du^k|y^k)\tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0)P_\pi(d\gamma^1, \ldots, d\gamma^N)\mathbb{P}_0(d\omega_0) \tag{B.4}$$
$$= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0),$$

where $\tilde{\mu}^N$ is the joint distribution of observations $(y^1,\dots,y^N)$, conditioned on ω_0. Equality (B.2) follows from (B.1), and (B.3) follows from relabeling $u^{\sigma(i)},y^{\sigma(i)}$ with u^i, y^i for all $i=1,\dots,N$. Equality (B.4) follows from Assumptions 1 and 7.

Let $\epsilon\geq 0$. Consider a randomized policy $P^*_{\pi,\epsilon}\in\bar{L}^N$ such that

$$\int P^*_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\leq \inf_{P_\pi\in\bar{L}^N}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)+\epsilon. \tag{B.5}$$

Consider $\tilde{P}_{\pi,\epsilon}$ as a convex combination of all possible permutations of $P^*_{\pi,\epsilon}$ by averaging them. Because $\bar{L}^N$ is convex, we have $\tilde{P}_{\pi,\epsilon}\in\bar{L}^N$. Also, we have that $\tilde{P}_{\pi,\epsilon}\in L^N_{\mathrm{EX}}$ because for any permutation $\sigma\in S_N$,

$$\begin{aligned}\tilde{P}_{\pi,\epsilon}(d\gamma^1,\dots,d\gamma^N)&:=\sum_{\sigma\in S_N}\frac{1}{|S_N|}P^{*,\sigma}_{\pi,\epsilon}(d\gamma^1,\dots,d\gamma^N)\\&=\tilde{P}^{\sigma}_{\pi,\epsilon}(d\gamma^1,\dots,d\gamma^N),\end{aligned}$$

where $|S_N|$ denotes the cardinality of the set S_N, and the second equality follows from the fact that the sum is over all permutations σ by taking average of them. Therefore, a randomized policy $\tilde{P}_{\pi,\epsilon}$ is in $\bar{L}^N\cap L^N_{\mathrm{EX}}$. We have

$$\begin{aligned}\int\tilde{P}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)&:=\int\left(\sum_{\sigma\in S_N}\frac{1}{|S_N|}P^{*,\sigma}_{\pi,\epsilon}\right)(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\&=\sum_{\sigma\in S_N}\frac{1}{|S_N|}\int P^{*,\sigma}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\&=\sum_{\sigma\in S_N}\frac{1}{|S_N|}\int P^{*}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\\&\leq\inf_{P_\pi\in\bar{L}^N}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)+\epsilon,\end{aligned}$$

where the second equality is true because the map $P_\pi\longmapsto\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)$ is linear. The third equality follows from (B.4), and the inequality follows from (B.5). Because $\tilde{P}_{\pi,\epsilon}\in\bar{L}^N\cap L^N_{\mathrm{EX}}$, we have

$$\int\tilde{P}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\geq\inf_{P_\pi\in\bar{L}^N\cap L^N_{\mathrm{EX}}}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0).$$

Hence, we have

$$\inf_{P_\pi\in\bar{L}^N\cap L^N_{\mathrm{EX}}}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\leq\inf_{P_\pi\in\bar{L}^N}\int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)+\epsilon.$$

This completes the proof because ϵ is arbitrary. □

B.2. Proof of Lemma 2

To prove Lemma 2, we use two following results by Diaconis and Friedman [33, theorem 13] and Aldous et al. [1, proposition 7.20] (see also Kallenberg [49] for more general results), which we recall for reader's convenience.

Theorem B.1 (Diaconis and Freedman [33, Theorem 13]). *Let $Y=(Y_1,\dots,Y_n)$ be an n-exchangeable and $Z=(Z_1,Z_2,\dots)$ an infinitely exchangeable sequence of random variables with $\mathcal{L}(Z_1,\dots,Z_k)=\mathcal{L}(Y_{I_1},\dots,Y_{I_k})$ for all $k\geq 1$, where the indices $(I_1,I_2,\dots)$ are i.i.d. random variables with the uniform distribution on the set $\{1,\dots,n\}$. Then, for all $m=1,\dots,n$,*

$$\|\mathcal{L}(Y_1,\dots,Y_m)-\mathcal{L}(Z_1,\dots,Z_m)\|_{TV}\leq\frac{m(m-1)}{2n}, \tag{B.6}$$

where $\mathcal{L}(\cdot)$ denotes the law of random variables, and $\|\cdot\|_{TV}$ is the total variation norm.

Theorem B.2 (Aldous et al. [1, Proposition 7.20]). *Let $X:=(X_1,X_2,\dots)$ be an infinitely exchangeable sequence of random variables taking values in a Polish space $\mathbb{X}$ and directed by a random measure α (i.e., α is a $\mathcal{P}(\mathbb{X})$-valued random variable, and $Pr(X\in A)=\int_{\mathcal{P}(\mathbb{X})}\prod_{i=1}^{\infty}\xi(A^i)Pr(\alpha\in d\xi)$, where $A^i\in\mathcal{B}(\mathbb{X})$ and $(A=A^1\times A^2\times\dots)$; see Aldous et al. [1, definition 2.6]). Suppose that, for each n, either*

1. *$X^{(n)}=(X^{(n)}_1,X^{(n)}_2,\dots)$ is infinitely exchangeable directed by α_n or*
2. *$X^{(n)}=(X^{(n)}_1,\dots,X^{(n)}_n)$ is n-exchangeable with empirical measure α_n.*

Then, $X^{(n)}$ converges in distribution to X ($X^{(n)}\xrightarrow[n\to\infty]{d}X$) if and only if $\alpha_n\xrightarrow[n\to\infty]{d}\alpha$.

We note that by convergence in distribution to an infinite exchangeable sequence, we mean the following: $X^{(n)} \xrightarrow[n\to\infty]{d} X$ if and only if $(X_1^{(n)},\dots,X_m^{(n)}) \xrightarrow[n\to\infty]{d} (X_1,\dots,X_m)$ for each $m \geq 1$ (Aldous et al. [1, p. 55]).

Using the above theorems, we now complete the proof of Lemma 2. Because the action space $\mathbb{U}$ is compact and observations are i.i.d. with a fixed marginal (under Assumption 2, via a change of measure argument, observations can be viewed to be independent of ω_0), the set of probability measures L^N is tight. Furthermore, by Yüksel [83, theorem 5.1], L^N is closed under the topology of weak convergence, and hence L^N is compact. Using the argument in Yüksel [83, theorem 5.1] under Assumption 4, the expected cost function is lower semicontinuous in policies $P_N \in L^N$. Hence, there exists an optimal policy for $(\mathcal{P}_N)$, and by Lemma 1, this optimal policy can be assumed to be in L^N_{EX}. Consider a sequence of N-exchangeable randomized policies $\{P^{*,N}_\pi\}_N$, where for every $N \geq 1$, $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$ and

$$\int P^{*,N}_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) = \inf_{P^N_\pi \in L^N_{\mathrm{EX}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0). \tag{B.7}$$

In the following, we show (22) in two steps. In the first step, for every N, we use Theorem B.1 to construct an infinitely exchangeable randomized policy $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}$ using the N-exchangeable randomized policy $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$, by considering the indices as a sequence of i.i.d. random variables with uniform distribution on the set $\{1,\dots,N\}$, and then we show that there exists a subsequence of joint measures, converging weakly in the first coordinate, observations, and the average of induced actions of randomized policies $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}$. In the second step, we show that the expected cost function induced by the N-exchangeable randomized policy $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$ converges through a subsequence to a limit induced by an infinitely exchangeable randomized policy $P^{*,\infty}_{\pi,N}$.

Step 1. Let $(I_1,I_2,\dots)$ be i.i.d. random variables with uniform distribution on the set $\{1,\dots,N\}$. For a fixed N and for any N-exchangeable randomized policy $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$, we construct an infinitely exchangeable randomized policy $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}$ as follows: for every N and m and for all $A^i \in \mathcal{B}(\Gamma^i)$,

$$P^{*,\infty}_{\pi,N}(\gamma^1 \in A^1,\dots,\gamma^m \in A^m) := P^{*,N}_\pi(\gamma^{I_1} \in A^1,\dots,\gamma^{I_m} \in A^m).$$

where $P^{*,\infty}_{\pi,N}$ is the restriction of $P^{*,\infty}_{\pi,P^N_\pi} \in L_{\mathrm{EX}}$ to the first N components. We note that $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}$ because we use i.i.d. sequence $(I_1,I_2,\dots)$ for indexing probability measures on the space of policies. Hence, for every fixed N and N-exchangeable randomized policy $P^{*,N}_\pi$, a randomized policy $P^{*,\infty}_{\pi,N}$ is i.i.d across DMs, and hence, it is infinitely exchangeable.

Let $u^{*,i}_N$ be the control action induced by γ^i_N, where random variables $(\gamma^1_N,\dots,\gamma^N_N)$ are determined by N-exchangeable randomized policy $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$. Let $u^{*,i}_{\infty,N}$ be the control action induced by $\gamma^i_{N,\infty}$, where random variables $(\gamma^1_{N,\infty},\dots,\gamma^N_{N,\infty})$ are determined by infinitely exchangeable randomized policy $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}$. Because under the reduction (Assumption 2), observations are i.i.d. and also independent of ω_0, following from Theorem B.1, we have, for every $m \geq 1$,

$$\begin{aligned}&\left\|\mathcal{L}(\gamma^1_N,\dots,\gamma^m_N,y^1,\dots,y^m) - \mathcal{L}(\gamma^1_{N,\infty},\dots,\gamma^m_{N,\infty},y^1,\dots,y^m)\right\| \\ &= \left\|\mathcal{L}(\gamma^1_N,\dots,\gamma^m_N)\prod_{i=1}^m \mathcal{L}(y^i) - \mathcal{L}(\gamma^1_{N,\infty},\dots,\gamma^m_{N,\infty})\prod_{i=1}^m \mathcal{L}(y^i)\right\|_{TV} \xrightarrow[N\to\infty]{} 0,\end{aligned} \tag{B.8}$$

where (B.8) follows from the fact that $(\gamma^1_N,\dots,\gamma^N_N)$ and $(\gamma^1_{N,\infty},\dots,\gamma^N_{N,\infty})$ are random variables with joint probability measures $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$ and $P^{*,\infty}_{\pi,N} \in L_{\mathrm{EX}}|_N$, respectively. The set $L_{\mathrm{EX}}|_N$ corresponds to the set of N-DM randomized policies that are the restrictions of policies in L_{EX} to the N first components.

Because $\mathbb{U}$ is compact, the marginal of probability measures on $\mathbb{U}$ is tight. Because the probability measure on $\mathbb{Y}$ is fixed, the marginal on $\mathbb{Y}$ is also tight. Because marginals are tight, the collection of all measures on $\mathbb{U}\times\mathbb{Y}$ with these tight marginals is also tight (see, e.g., Yüksel [82, proof of theorem 2.4]), and hence, the set Γ^i is tight for each $i \in \mathbb{N}$. Hence, $\{\mathcal{L}(\gamma^i_{\infty,N})\}_N$ is tight for each DM, and by exchangeablity, $\mathcal{L}(\gamma^i_{\infty,N}) = \mathcal{L}(\gamma^1_{\infty,N})$ for all $i \in \mathbb{N}$. Hence, there exists a subsequence (denoted by the index l) such that $\mathcal{L}(\gamma^i_{\infty,l}) \underset{l\to\infty}{\Rightarrow} \mathcal{L}(\gamma^i_\infty)$ for all $i \in \mathbb{N}$. Because marginals of $\{\mathcal{L}(\gamma^1_{\infty,l},\dots,\gamma^m_{\infty,l})\}_l$ are tight, for each $m \geq 1$, there exists a further subsequence denoted by index n such that

$$\mathcal{L}(\gamma^1_{\infty,n},\dots,\gamma^m_{\infty,n}) \underset{n\to\infty}{\Longrightarrow} \mathcal{L}(\gamma^1_\infty,\dots,\gamma^m_\infty),$$

where $(\gamma^1_\infty,\gamma^2_\infty,\dots)$ is infinitely exchangeable and induced by an infinitely exchangeable randomized policy $P^{*,\infty}_\pi \in L_{\mathrm{EX}}$ because L_{EX} is closed under the weak convergence topology, where by weak convergence for an infinite sequence, we mean weak convergence of finite restrictions. That is because, if $P^{\sigma,*,\infty}_\pi$ is the limit in the weak convergence topology of the sequence of randomized policies $\{P^{\sigma,*,\infty}_{\pi,n}\}_n$ as $n \to \infty$, where for $A^i \in \mathcal{B}(\Gamma^i)$ and for all $N \in \mathbb{N}$ and all finite permutations

$\sigma \in S_N$,

$$P^{\sigma,*,\infty}_{\pi,n}(\gamma^1 \in A^1, \gamma^2 \in A^2, \dots) := P^{*,\infty}_{\pi,n}(\gamma^{\sigma(1)} \in A^1, \gamma^{\sigma(2)} \in A^2, \dots),$$

then, following from exchangeability, because sequences $\{P^{*,\infty}_{\pi,n}\}_n$ and $\{P^{\sigma,*,\infty}_{\pi,n}\}_n$ are identical, the limits in the weak convergence topology of both randomized policies $P^{*,\infty}_{\pi}$ and $P^{\sigma,*,\infty}_{\pi}$ are also identical, and hence, the limit $P^{*,\infty}_{\pi}$ is infinitely exchangeable, $P^{*,\infty}_{\pi} \in L_{\mathrm{EX}}$. Hence, following from (B.8), for each $m \geq 1$,

$$\mathcal{L}(\gamma^1_n, \dots, \gamma^m_n) \underset{n\to\infty}{\longrightarrow} \mathcal{L}(\gamma^1_\infty, \dots, \gamma^m_\infty).$$

By construction of random variables $u^{*,i}_n$ and $u^{*,i}_\infty$ and because random variables γ^i_n are independent of y^is, we have, for each $m \geq 1$,

$$(u^{*,1}_n, \dots, u^{*,m}_n) \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} (u^1_\infty, \dots, u^m_\infty),$$

where $(u^1_\infty, u^2_\infty, \dots)$ is induced by an infinitely exchangeable policy $P^{*,\infty}_{\pi} \in L_{\mathrm{EX}}$. Following from Theorem B.2, we have, for all $A \in \mathcal{U}$ and $\mathbb{P}$-almost surely,

$$F_n(A) := F^\omega_n(A) := \frac{1}{n}\sum_{i=1}^n \delta_{u^{*,i}_n(\omega)}(A) \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} \alpha^\omega(A), \tag{B.9}$$

where ω denotes the sample path dependency, and α is the directing measure of infinitely exchangeable random variables $(u^1_\infty, u^2_\infty, \dots)$ (that is, $\alpha(\omega, A) = Pr(u^{*,i}_\infty \in A | H)$ almost surely for all $A \in \mathcal{U}$, where H is the σ-field generated by $\mathcal{P}(\mathbb{U})$-valued random variable α; Aldous et al. [1]). Following from (B.9), because the action space $\mathbb{U}$ is compact, we have, $\mathbb{P}$-almost surely,

$$\bar{F}_n := \bar{F}^\omega_n := \frac{1}{n}\sum_{i=1}^n u^{*,i}_n(\omega) = \int_{\mathbb{U}} u F_n(du) \underset{n\to\infty}{\overset{\mathrm{d}}{\longrightarrow}} \bar{F} := \int_{\mathbb{U}} u \alpha^\omega(du). \tag{B.10}$$

Define $\tilde{P}^{*,n}$ as the joint probability measure of $(u^{*,1}_n, \bar{F}_n, \underline{y})$ where marginals on $\underline{y} := (y^1, y^2, \dots)$ are fixed to be $\prod_{i=1}^\infty Q(dy^i)$. Because marginals on $(u^{*,1}_n, \bar{F}_n)$ are tight and marginals on $\underline{y}$ are fixed, $\{\tilde{P}^{*,n}\}_n$ is tight. Hence, there exists a sub-subsequence $\{\tilde{P}^{*,k}\}_k$ that converges weakly to $\tilde{P}^*$ as k goes to infinity. This implies that the marginals $\{\tilde{P}^{*,k}\}_k$ on $(u^{*,1}_k, \bar{F}_k)$ converge to the marginals of $\tilde{P}^*$ on $(u^{*,1}, \bar{F})$, and hence, $\tilde{P}^*$ is induced by $(u^1_\infty, u^2_\infty, \dots)$, which is infinitely exchangeable and is induced by an infinitely exchangeable randomized policy in L_{EX}.

Step 2. We have

$$\begin{aligned}
&\limsup_{N\to\infty} \int P^{*,N}_{\pi}(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \int c\left(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p\right) \prod_{k=1}^N \gamma^k(du^k | y^k) P^{*,N}_{\pi}(d\gamma^1, \dots, d\gamma^N) \prod_{i=1}^N \hat{\mu}(dy^i | \omega_0) \mathbb{P}_0(d\omega_0) \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \int c\left(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p\right) \prod_{k=1}^N \gamma^k(du^k | y^k) P^{*,N}_{\pi}(d\gamma^1, \dots, d\gamma^N) \prod_{i=1}^N f(\omega_0, y^i) Q(dy^i) \mathbb{P}_0(d\omega_0) \qquad \text{(B.11)} \\
&= \limsup_{N\to\infty} \int\int_{\prod_{i=N+1}^\infty \mathbb{Y}} c(\omega_0, u^1, \bar{F}_N) \tilde{P}^{*,N}(du^1, d\bar{F}_N, d\underline{y}) \prod_{i=1}^\infty f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \qquad \text{(B.12)} \\
&\geq \lim_{k\to\infty} \int\int_{\prod_{i=k+1}^\infty \mathbb{Y}} c(\omega_0, u^1, \bar{F}_k) \tilde{P}^{*,k}(du^1, d\bar{F}_k, d\underline{y}) \prod_{i=1}^\infty f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \qquad \text{(B.13)} \\
&= \int c(\omega_0, u^1, \bar{F}) \tilde{P}^*(du^1, d\bar{F}, d\underline{y}) \prod_{i=1}^\infty f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \qquad \text{(B.14)} \\
&\geq \limsup_{N\to\infty} \inf_{P_\pi \in L_{\mathrm{EX}}} \int P_{\pi,N}(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0), \qquad \text{(B.15)}
\end{aligned}$$

where $\hat{\mu}$ is the conditional distribution of each observation y^i given ω_0. Under Assumption 3i and Assumption 2, using a change of measure argument as in (6), we can rewrite the expected cost function with respect to $\mathbb{P}$ equivalently as a new

expected cost function $c\left(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^{N} u^p\right)\prod_{i=1}^{N} f(\omega_0, y^i)$ with respect to a reference measure Q, under which observations are i.i.d. and independent of ω_0. Hence, (B.11) follows from Assumption 3i and Assumption 2, and (B.12) follows from integrating over the set $\prod_{i=N+1}^{\infty}\mathbb{Y}$ and because $(u_N^{*,1},\dots,u_N^{*,N})$ is N-exchangeable. Inequality (B.13) follows from the assumption that the cost function is bounded and limsup is the greatest subsequential limit where k is the index of the subsequence considered in Step 1. Equality (B.14) follows from the dominated convergence theorem (by Assumption 4) and by Step 1 because $\{\tilde{P}^{*k}\}_k$ converges weakly to $\tilde{P}^*$ as k goes to infinity. Inequality (B.15) follows from the fact that $\tilde{P}^*$ is the joint measure with the first coordinate $(u_\infty^1, u_\infty^2,\dots)$ that is infinitely exchangeable, and it is induced by an infinitely exchangeable randomized policy in L_{EX}. The above inequalities become equalities because the opposite direction is true as well (i.e., because $L_{\mathrm{EX}}|_N \subset L_{\mathrm{EX}}^N$). This completes the proof. □

B.3. Proof of Theorem 2

We complete the proof in four steps.

Step 1. Similar to the proof of Lemma 2, using Yüksel [83, theorem 5.1], we can show that there exists a randomized optimal policy for $(\mathcal{P}_N)$, belonging to the set L^N, and by Lemma 1, this randomized optimal policy can be assumed to be in the set of N-exchangeable randomized policies L_{EX}^N. Consider a sequence of N-exchangeable randomized policies $\{P_\pi^{*,N}\}_N$, where, for every $N \geq 1$, $P_\pi^{*,N} \in L_{\mathrm{EX}}^N$ and satisfies (B.7).

Step 2. In this step, we show that to establish an existence result, it is sufficient to show the convergence of the expected cost induced by a randomized optimal policy in $L_{\mathrm{PR,SYM}}^N$ of N-DM teams to the expected cost induced by a randomized policy $L_{\mathrm{PR,SYM}}$ of mean-field teams through a subsequence as N goes to infinity. We first lift the space of randomized admissible policies, and we represent any admissible randomized policy as a probability measure in L (which is convex) and $L_{\mathrm{EX}} \subset L$. We have

$$\inf_{P_\pi\in L}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)$$

$$\geq \limsup_{N\to\infty}\inf_{P_\pi^N\in L^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \tag{B.16}$$

$$= \limsup_{N\to\infty}\inf_{P_\pi^N\in L_{\mathrm{EX}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \tag{B.17}$$

$$\geq \lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P_\pi^N\in L_{\mathrm{EX}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\} \tag{B.18}$$

$$= \lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P_\pi\in L_{\mathrm{EX}}}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\} \tag{B.19}$$

$$= \lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P_\pi^N\in L_{\mathrm{CO,SYM}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\} \tag{B.20}$$

$$= \lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P_\pi^N\in L_{\mathrm{PR,SYM}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\} \tag{B.21}$$

$$\geq \inf_{P_\pi\in L_{\mathrm{PR,SYM}}}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \tag{B.22}$$

$$\geq \inf_{P_\pi\in L_{\mathrm{CO,SYM}}}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \tag{B.23}$$

$$\geq \inf_{P_\pi\in L}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0), \tag{B.24}$$

where (B.16) follows from exchanging limsup with inf and the fact that $P_{\pi,N} \in L^N$ is the restriction to N-first coordinate for any randomized policy $P_\pi \in L$, and (B.17) follows from Lemma 1. Inequality (B.18) follows from $\min\{M, c^N(\underline{\gamma},\underline{y},\omega_0)\} \leq c^N(\underline{\gamma},\underline{y},\omega_0)$. Equality (B.19) follows from Lemma 2, and (B.20) follows from Theorem 1. The set of extreme points of the convex set $L_{\mathrm{CO,SYM}}^N$ is in $L_{\mathrm{PR,SYM}}^N$ (that is because $L_{\mathrm{CO,SYM}}^N$ corresponds to the randomized policies with common and individual independent randomness, where each DM selects an identical randomized policy); hence, (B.21) is true because

$L^N_{\mathrm{CO,SYM}}$ is convex, and the map $\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) : L^N_{\mathrm{CO,SYM}} \to \mathbb{R}$ is linear. Inequalities (B.23) and (B.24) follow from the fact that $L_{\mathrm{PR,SYM}} \subset L_{\mathrm{CO,SYM}} \subset L$. Hence, by (B.24), this chain of inequalities must be a chain of equalities.

In the next two steps, we justify (B.22) through showing that there exists a subsequence of policies induced by symmetric/identical private randomization whose weak limit achieves (B.22). First, we establish compactness of the set of randomized policies $L^N_{\mathrm{PR,SYM}}$, and then we show a lower semicontinuity of the induced expected cost that justifies (B.22).

Step 3. Consider the set of randomized policies $L^N_{\mathrm{PR,SYM}}$. For each DM, we can equivalently represent any randomized policy as a probability measure on $\mathbb{U}\times\mathbb{Y}$, where the marginal on observations is fixed. Because the team is static, this decouples the policy spaces from the policies of the previous DMs. Following from symmetry, we can represent each DM's privately randomized policy space as $\{P \in \mathcal{P}(\mathbb{U}\times\mathbb{Y}) | P(B) = \int_B \Pi(du^i|y^i)Q(dy^i)\}$, where $B \in \mathcal{B}(\mathbb{U}\times\mathbb{Y})$, and Π is an identical randomized policy from the set of stochastic kernels from the space of observations to the space of actions for each DM. Because $\mathbb{U}$ is compact, the marginals on $\mathbb{U}$ are tight. Because the marginals are tight, the collection of all measures with these tight marginals is also tight (see, e.g., Yüksel [82, proof of theorem 2.4]), and, hence, the randomized policy space is tight. Following from symmetry, the set of individual randomized policies for each DM is closed under product topology where each coordinate converges in the weak convergence topology. Hence, this step concludes that there exists a subsequence of (symmetric) individually randomized policies for each DM that converges weakly to the limit that is identical for each DM. In Step 4, we show that the limit randomized policy is optimal by showing a lower semicontinuity of the induced expected cost.

Step 4. Define the empirical measure on actions and observations as follows:

$$\Lambda_N(B) := \frac{1}{N}\sum_{i=1}^N \delta_{\beta^i_N}(B), \tag{B.25}$$

where for each N, $\beta^i_N := (u^{i,*}_N, y^i)$, $B \in \mathcal{Z} := \mathbb{U}\times\mathbb{Y}$, and $u^{i,*}_N$ is the action induced by the randomized policy Π^*_N of Step 3.

Now, we have

$$\lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P^N_\pi\in L^N_{\mathrm{PR,SYM}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\}$$

$$= \lim_{M\to\infty}\limsup_{N\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_N(du\times\mathbb{Y})\right)\right\}\Lambda_N(du,dy)\right)\prod_{i=1}^\infty P^{*,\omega_0}_N(du^i,dy^i)\mathbb{P}_0(d\omega_0) \tag{B.26}$$

$$\geq \lim_{M\to\infty}\lim_{n\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_n(du\times\mathbb{Y})\right)\right\}\Lambda_n(du,dy)\right)\prod_{i=1}^\infty P^{*,\omega_0}_n(du^i,dy^i)\mathbb{P}_0(d\omega_0) \tag{B.27}$$

$$= \lim_{M\to\infty}\int\lim_{n\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_n(du\times\mathbb{Y})\right)\right\}\Lambda_n(du,dy)\right)\prod_{i=1}^\infty P^{*,\omega_0}_n(du^i,dy^i)\mathbb{P}_0(d\omega_0) \tag{B.28}$$

$$\geq \lim_{M\to\infty}\int\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda(du\times\mathbb{Y})\right)\right\}\Lambda(du,dy)\right)\prod_{i=1}^\infty P^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0) \tag{B.29}$$

$$= \limsup_{N\to\infty}\int\frac{1}{N}\sum_{i=1}^N c\left(\omega_0,u^i,\frac{1}{N}\sum_{p=1}^N u^p\right)\prod_{i=1}^N P^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0) \tag{B.30}$$

$$\geq \inf_{P_\pi\in L_{\mathrm{PR,SYM}}}\limsup_{N\to N}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0), \tag{B.31}$$

where $P^{*,\omega_0}_N(du^i,dy^i) := \Pi^*_N(du^i|dy^i)\hat{\mu}(dy^i|\omega_0) = \Pi^*_N(du^i|y^i)f(\omega_0,y^i)Q(dy^i)$. Equality (B.26) follows from (B.25), Assumption 3i, and symmetry of the optimal policies (because every DM applies an identical policy), and because the set of policies can be extended to an infinite product space by considering the expected cost by integrating over $\prod^\infty_{i=N}(\mathbb{U}\times\mathbb{Y})$. Inequality (B.27) follows from the fact that limsup is the greatest convergent subsequential limit for a bounded sequence, where we denoted the convergent subsequence of coordinates of policies in Step 3 by the index $n \in \mathbb{I} \subset \mathbb{N}$. Equality (B.28) follows from the law of total expectation and the dominated convergence theorem.

Fix the convergent subsequence indexed by n. Following from symmetry and Assumptions 2 and 3i, we have that $\beta^i_n = (u^{*,i}_n, y^i)$ are i.i.d. Now, using an argument similar to the one used in Sanjari and Yüksel [70, proof of theorem 8], through choosing a suitable sub-subsequence and using the strong law of large numbers, we can show that for any continuous

bounded function $g \in C_b(\mathcal{Z})$,

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} g(\beta_n^i) - \mathbb{E}(g(\beta_\infty^i))\right| = 0\right\}\right) = 1. \tag{B.32}$$

By considering a countable family of measure-determining functions $\mathcal{T} \subset C_b(\mathcal{Z})$, (B.32) implies that the empirical measures $\{\Lambda_n\}_n$ converge weakly to $\Lambda = \mathcal{L}(\beta_\infty^i)$ $\mathbb{P}$-almost surely, and Λ is induced by the limiting randomized policy P^{*,ω_0}. We equip the above set of probability measures on $\Omega_0 \times \mathbb{U} \times \mathbb{Y}$ with the w-s topology, that is, the coarsest topology on $\mathcal{P}(\Omega_0 \times \mathbb{U} \times \mathbb{Y})$ under which $\int \hat{f}(\omega_0, u, y)\kappa(d\omega_0, du, dy) : \mathcal{P}(\Omega_0 \times \mathbb{U} \times \mathbb{Y}) \to \mathbb{R}$ is continuous for every measurable and bounded $\hat{f}$ that is continuous in u but need not be continuous in y and ω_0 (see, e.g., Schäl [72] and Yüksel [83, theorem 5.6]). Following from Assumption 5 and Assumption 3ii, and because actions induced by identical policies are i.i.d. (thanks to symmetry), we have, $\mathbb{P}$-almost surely,

$$f_n := \min\left\{M, c\left(\omega_0, \cdot, \int_{\mathbb{U}} u\Lambda_n(du \times \mathbb{Y})\right)\right\} \xrightarrow{\text{cont}} f := \min\left\{M, c\left(\omega_0, \cdot, \int_{\mathbb{U}} u\Lambda(du \times \mathbb{Y})\right)\right\},$$

where we recall that the sequence f_n converges continuously to f ($f_n \xrightarrow{\text{cont}} f$) if and only if $f_n(a_n) \to f(a)$ whenever $a_n \to a$ as $n \to \infty$. Now, (B.29) follows from the generalized dominated convergence theorem for varying measures in Serfozo [74, theorem 3.5]. Equality (B.30) follows from the monotone convergence theorem and an analysis similar to that established in (B.29) using the fact that P^{*,ω_0} does not depend on N and is symmetric across DMs. Inequality (B.31) follows from the fact that $P^{*,\omega_0}(du^i, dy^i) := \Pi^*(du^i|y^i)f(\omega_0, y^i)Q(dy^i)$, achieving (B.30), belongs to $L_{\text{PR,SYM}}$. That is because, following from Step 3, for each DM, the set of randomized policies is closed under the product topology, where each coordinate converges weakly, and hence, the limiting policy is also a randomized policy induced by a subsequence of N-DM optimal policies (which are symmetric across DMs). This implies that (B.31) holds, which implies (B.22) and completes the proof. □

Appendix C. Proofs from Section 5

C.1. Independent Measurement Reduction Under Assumption 8

Under Assumption 8i, we can represent the expected cost for deterministic policies as

$$\begin{aligned}
J_N(\underline{\gamma}^{1:N}) &:= \int c(\omega_0, u_{0:T-1}^1, \ldots, u_{0:T-1}^N)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\
&\quad \times \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma_t^i(y_t^i)\in du_t^i\}} \nu_t^i\left(dy_t^i \,|\, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right) \\
&= \int c(\omega_0, u_{0:T-1}^1, \ldots, u_{0:T-1}^N)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\
&\quad \times \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma_t^i(y_t^i)\in du_t^i\}} \psi_t^i\left(y_t^i, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right)\tau_t^i(dy_t^i) \\
&= \int c_s\left(\omega_0, \underline{\zeta}^{1:N}, u_{0:T-1}^{1:N}, y_{0:T-1}^{1:N}\right)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma_t^i(y_t^i)\in du_t^i\}}\tau_t^i(dy_t^i),
\end{aligned} \tag{C.1}$$

where the new (equivalent) cost function is defined as

$$c_s\left(\omega_0, \underline{\zeta}^{1:N}, u_{0:T-1}^{1:N}, y_{0:T-1}^{1:N}\right) := c\left(\omega_0, u_{0:T-1}^{1:N}\right)\prod_{i=1}^{N}\prod_{t=0}^{T-1}\psi_t^i\left(y_t^i, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\right),$$

and (C.1) follows from Assumption 8i. Similarly, we can define the new (equivalent) cost function under Assumption 8ii. We note that in the above, we considered control actions induced by deterministic policies; however, the above analysis can be extended to randomized policies by just replacing $\prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma_t^i(y_t^i)\in du_t^i\}}$ with $\prod_{i=1}^{N}\prod_{t=0}^{T-1}\gamma_t^i(du_t^i|y_t^i)$.

C.2. Proof of Lemma 3

We follow the steps of the proof of Lemma 1. For any permutation $\sigma \in S_N$, we define a randomized policy $P_\pi^\sigma \in \bar{L}^N$ as a permutation σ of arguments of a randomized policy $P_\pi \in \bar{L}^N$; that is, for $A^i \in \mathcal{B}(\Gamma^i)$,

$$P_\pi^\sigma(\underline{\gamma}^1 \in A^1, \ldots, \underline{\gamma}^2 \in A^N) := P_\pi(\underline{\gamma}^{\sigma(1)} \in A^1, \ldots, \underline{\gamma}^{\sigma(N)} \in A^N). \tag{C.2}$$

We have

$$\int P_\pi^\sigma(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$= \int c(\omega_0,\underline{u}^1,\dots,\underline{u}^N)\prod_{k=1}^N \underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)P_\pi(d\underline{\gamma}^{\sigma(1)},\dots,d\underline{\gamma}^{\sigma(N)})\tilde{\mu}^N(d\underline{\zeta}^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)$$

$$\times \prod_{t=0}^{T-1}\prod_{i=1}^N \nu_t^i\Big(dy_t^i|\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^{1:N},u_{t-1}^{1:N}\Big) \tag{C.3}$$

$$= \int c(\omega_0,\underline{u}^{\sigma(1)},\dots,\underline{u}^{\sigma(N)})\prod_{k=1}^N \underline{\gamma}^{\sigma(k)}(d\underline{u}^{\sigma(k)}|\underline{y}^{\sigma(k)})P_\pi(d\underline{\gamma}^1,\dots,d\underline{\gamma}^N)\tilde{\mu}^N(d(\underline{\zeta}^\sigma)^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)$$

$$\times \prod_{t=0}^{T-1}\prod_{i=1}^N \nu_t^i\Big(dy_t^{\sigma(i)}\Big|\omega_0,x_0^{\sigma(i)},\zeta_{0:t-1}^{\sigma(i)},(y_{0:t-1}^\sigma)^{1:N},(u_{0:t-1}^\sigma)^{1:N}\Big) \tag{C.4}$$

$$= \int c(\omega_0,\underline{u}^1,\dots,\underline{u}^N)\prod_{k=1}^N \underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)P_\pi(d\underline{\gamma}^1,\dots,d\underline{\gamma}^N)\tilde{\mu}^N(d\underline{\zeta}^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)$$

$$\times \prod_{t=0}^{T-1}\prod_{i=1}^N \nu_t^i\Big(dy_t^i\Big|\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^{1:N},u_{0:t-1}^{1:N}\Big)$$

$$= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0), \tag{C.5}$$

where $\tilde{\mu}^N$ is the conditional distribution of uncertainties $\underline{\zeta}^{1:N}$ given ω_0. Equality (C.3) follows from Assumption 10ii and (C.2), and (C.4) follows from relabeling $\underline{u}^{\sigma(i)},\underline{y}^{\sigma(i)},\underline{\zeta}^{\sigma(i)}$ with $\underline{u}^i,\underline{y}^i,\underline{\zeta}^i$ for all $i=1,\dots,N$ and the fact that $y_t^i = h_t(x_0^i, x_0^{-i}, \zeta_{0:t}^i, \zeta_{0:t}^{-i}, u_{0:t-1}^i, u_{0:t-1}^{-i})$. Equality (C.5) follows from Assumption 10i, Assumption 9, and the hypothesis that the information structure is symmetric. The rest of the proof follows from steps similar to those used in the proof of Lemma 1. □

C.3. Proof of Lemma 4

We follow steps similar to those used in the proof of Lemma 2. In the following, we present only a sketch of the proof (for the complete proof, please see Sanjari et al. [71, proof of lemma 5.3]).

Under Assumption 11 and Assumption 5, for every finite N, there exists an optimal policy in L_{EX}^N. Consider a sequence $\{P_\pi^{*,N}\}_N$, where, for every $N\geq 1$, $P_\pi^{*,N}\in L_{\mathrm{EX}}^N$ and

$$\int P_\pi^{*,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$

$$= \inf_{P_\pi^N\in L_{\mathrm{EX}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0). \tag{C.6}$$

Step 1. Suppose that random variables $u_{t,N}^{*,i}$ are induced by $\gamma_{t,N}^i$, where $(\gamma_{t,N}^1,\dots,\gamma_{t,N}^N)$ for all $t=0,\dots,T-1$ are determined by the randomized policy $P_\pi^{*,N}\in L_{\mathrm{EX}}^N$. Suppose further that random variables $u_{t,\infty,N}^{*,i}$ are induced by $\gamma_{t,\infty,N}^i$, where $(\gamma_{t,\infty,N}^1,\dots,\gamma_{t,\infty,N}^N)$ are determined by the randomized policy $P_{\pi,N}^{*,\infty}\in L_{\mathrm{EX}}$. Let $\underline{\gamma}_N^i := (\gamma_{0,N}^i,\dots,\gamma_{T-1,N}^i)$, $\underline{\gamma}_{N,\infty}^i := (\gamma_{0,\infty,N}^i,\dots,\gamma_{T-1,\infty,N}^i)$, $\underline{u}_N^i := (u_{0,N}^i,\dots,u_{T-1,N}^i)$ and $\underline{u}_{N,\infty}^i := (u_{0,\infty,N}^i,\dots,u_{T-1,\infty,N}^i)$ for each DM. Because, under Assumption 8, observations are i.i.d. across DMs and also independent of ω_0, similar to the proof of Lemma 2, we can show that there exist a subsequence (denoted by n) such that for each $m\geq 1$,

$$(\underline{u}_n^{*,1},\dots,\underline{u}_n^{*,m})\xrightarrow[n\to\infty]{\mathrm{d}}(\underline{u}_\infty^1,\dots,\underline{u}_\infty^m),$$

where $(\underline{u}_\infty^1,\underline{u}_\infty^2,\dots)$ is induced by an infinitely exchangeable randomized policy $P_\pi^{*,\infty}\in L_{\mathrm{EX}}$. Following from Theorem B.2, for all $A\in\mathcal{U}$, we have, $\mathbb{P}$-almost surely,

$$F_{n,t}(A) := F_{n,t}^\omega(A) := \frac{1}{n}\sum_{i=1}^n \delta_{u_{t,n}^{*,i}(\omega)}(A)\xrightarrow[n\to\infty]{\mathrm{d}}\alpha_t^{u,\omega}(A), \tag{C.7}$$

where α_t^u is the directing random measure of infinitely exchangeable random variables $(\underline{u}_{\infty,t}^1,\underline{u}_{\infty,t}^2,\dots)$. By (C.7), because the action space $\mathbb{U}$ is compact, we have, for all $t=0,\dots,T-1$ and $\mathbb{P}$-almost surely,

$$\mu_{n,t}^u := \mu_{n,t}^{u,\omega} := \frac{1}{n}\sum_{i=1}^n u_{t,n}^{*,i} = \int_{\mathbb{U}} uF_{n,t}(du)\xrightarrow[n\to\infty]{\mathrm{d}}\mu_t^u := \int_{\mathbb{U}} u\alpha_t^{u,\omega}(du). \tag{C.8}$$

Step 2. Let $x^{*,i}_{t,n}$ be the state of DM^i at time t under $u^{*,i}_{0:t-1,n} := (u^{*,i}_{0,n}, \dots, u^{*,i}_{t-1,n})$, evolving by the following dynamics:

$$x^{*,i}_{t+1,n} = f_t\left(x^{*,i}_{t,n}, u^{*,i}_{t,n}, \frac{1}{n}\sum_{p=1}^{n} x^{*,p}_{t,n}, \frac{1}{n}\sum_{p=1}^{n} u^{*,p}_{t,n}, w^i_t\right). \tag{C.9}$$

Let $t = 1$. Because initial states are conditionally i.i.d. by continuity of the function f_0 in actions and states, we have that $x^{*,i}_{1,n} \xrightarrow[n\to\infty]{\mathrm{d}} x^{*,i}_{1,\infty}$ for all DMs. Hence, $(x^{*,1}_{1,n}, \dots, x^{*,m}_{1,n})$ is tight, and for each $m \geq 1$, there exists a sub-subsequence $(x^{*,1}_{1,k}, \dots, x^{*,m}_{1,k}) \xrightarrow[k\to\infty]{\mathrm{d}} (x^{*,1}_{1,\infty}, \dots, x^{*,m}_{1,\infty})$. Following from Theorem B.2, because f_0 is bounded, we have, $\mathbb{P}$-almost surely,

$$\mu^x_{k,1} := \frac{1}{n}\sum_{i=1}^{n} x^{*,i}_{1,k} = \mu^{x,\omega}_{k,1} = \int_{\mathbb{X}} x d\left(\frac{1}{k}\sum_{i=1}^{k} \delta_{x^{*,i}_{1,k}}\right) \xrightarrow[k\to\infty]{\mathrm{d}} \mu^x_1 := \int_{\mathbb{X}} x \alpha^{x,\omega}_1(dx), \tag{C.10}$$

where α^x_1 is the directing measure for $(x^{*,1}_{1,\infty}, x^{*,2}_{1,\infty}, \dots)$. By induction, for each $m \geq 1$, there exists a further sub-subsequence n (which we indicate by n to omit a further sub-sub-script) such that $(\underline{x}^{*,1}_n, \dots, \underline{x}^{*,m}_n) \xrightarrow[n_l\to\infty]{\mathrm{d}} (\underline{x}^{*,1}_\infty, \dots, \underline{x}^{*,m}_\infty)$ and $\mu^x_{n_l,t} \xrightarrow[n\to\infty]{\mathrm{d}} \mu^x_t$ for all $t = 0, \dots, T-1$.

The rest of the proof is similar to (Step 2 of) the proof of Lemma 2 (see Sanjari et al. [71, proof of lemma 5.3]) with the difference that in addition to actions and observations, we need to consider states and disturbances in our analysis. □

C.4. Proof of Theorem 3

We complete the proof using steps similar to those used in the proof of Theorem 2. In the following, we present only a sketch of the proof (for the complete proof, please see Sanjari et al. [71, proof of theorem 5.4]).

Step 1. Under Assumptions 5 and 11, by Lemma 3, for every finite N, there exists a randomized optimal policy in L^N_{EX}. Consider a sequence $\{P^{*,N}_\pi\}_N$, where for every $N \geq 1$, $P^{*,N}_\pi \in L^N_{\mathrm{EX}}$ and satisfies (C.6).

Step 2. Similar to Step 2 of the proof of Theorem 2 using Lemma 4 and Theorem 1, to complete the proof, it is sufficient to show that

$$\begin{aligned}
&\lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\mathrm{PR,SYM}}} \int P^N_\pi(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{\zeta}) \min\{M, c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\} \nu^N(d\underline{y} | \underline{\zeta}, \underline{\gamma}, \omega_0) \\
&\geq \inf_{P_\pi \in L_{\mathrm{PR,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{\zeta}) c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) \nu^N(d\underline{y} | \underline{\zeta}, \underline{\gamma}, \omega_0).
\end{aligned} \tag{C.11}$$

In the next two steps, we justify (C.11) through showing that there exists a subsequence of randomized policies induced by symmetric/identical private randomization whose weak subsequential limit achieves the right-hand side of (C.11).

Step 3. Consider the set of randomized policies $L^N_{\mathrm{PR,SYM}}$. Under a symmetric information structure and Assumption 8, and because each DM applies an identical policy, $\underline{y}^i$ are i.i.d. across DMs and also independent of ω_0. Hence, following from the information structure, the randomized policy spaces of each DM can be separated from the policies of the other DMs. Thus, we can equivalently represent any privately randomized policy for each DM acting through time separately as probability measures induced by symmetric (identical randomized) policies, that is, as probability measures q on $\mathbf{U} \times \mathbf{Y}$, where randomized policies for each DM and for every $t = 0, \dots, T-1$ satisfy the following equality:

$$\begin{aligned}
&\int g\left(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^i_{0:t}\right) q\left(dy^i_{0:t}, du^i_{0:t} | \omega_0, x^i_0, \zeta^i_{0:t-1}\right) \\
&= \int g\left(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^i_{0:t}\right) \prod_{k=0}^{t} \Pi^N_k(du^i_k | y^i_k) \eta_k\left(dy^i_k | \omega_0, x^i_0, \zeta^i_{0:k-1}, y^i_{0:k-1}, u^i_{0:k-1}\right),
\end{aligned}$$

for all bounded functions g continuous in actions and observations and measurable in other arguments, and for some stochastic kernels Π^N_k, representing a randomized policy of DMs at time k (which is identical across DMs).

Because $\mathbb{U}$ is compact, the marginals on $\mathbf{U}$ are tight under the weak convergence topology. Hence, the collection of all probability measures with these tight marginals is also tight (see, e.g., Yüksel [82, proof of theorem 2.4]). Because every DM applies an identical policy and because observations are i.i.d., the randomized policy space is tight, and hence, there exists a subsequence of randomized policies $\{\tilde{q}_n\}_n \subseteq \mathcal{P}(\prod_t(\mathbf{Y} \times \mathbf{U}))$ that converges weakly (each coordinate converges weakly) to a limit $\tilde{q}$ (as an infinite product of policies of DMs), where n is the index of the subsequence, and n goes to infinity. Now, we show that the randomized policy spaces are closed under the weak convergence topology. Suppose that $\{\hat{q}_n\}_n \subseteq \mathcal{P}(\mathbf{Y} \times \mathbf{U})$ (induced by identical randomized policies Π^n_t for each DM at time $t = 0, \dots, T-1$) converges weakly to $\hat{q}$. If Assumption 8i (under the structure of Assumption 11) holds, then there exists an independent static reduction for each DM over time, and hence, following from the discussion in the proof of Yüksel [83, theorem 5.2], each coordinate of policy spaces corresponds to DM^i at time t is closed under the weak convergence topology. Also, if Assumption 8ii (under the structure of Assumption 11) holds, then Yüksel [83, theorem 5.6] leads to the same conclusion. Hence, this

implies that for $\{\tilde{q}^*_N\}_N \subseteq \mathcal{P}(\prod_{i=1}^N(\mathbf{Y}\times\mathbf{U}))$, induced by optimal randomized policies $\Pi^{*,N}_t$ for each DM at time t, there exists a subsequence $\{\tilde{q}^*_n\}_n \subseteq \mathcal{P}(\prod_{i=1}^\infty(\mathbf{Y}\times\mathbf{U}))$ (as an infinite product of policies $\Pi^{*,n}_t$) that converges weakly (each coordinate converges weakly) to a limit $\tilde{q}^*$. This limit belongs to $L_{\mathrm{PR,SYM}}$ and is induced by a randomized policy $\Pi^{*,\infty}_t$ for each DM at time t.

Step 4. An argument similar to the one used in Step 4 of the proof of Theorem 2 can be used to show that (C.11) holds, and this completes the proof. □

Appendix D. Proof from Section 6

D.1. Proof of Theorem 4

To prove part i, we first show that (28) holds. We have

$$\inf_{P^N_\pi\in L^N_{\mathrm{CO}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \geq \inf_{P^N_\pi\in L^N_{\mathrm{CO}}\cap L_{\mathrm{EX}}|_N}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)-\epsilon_N \tag{D.1}$$

$$= \inf_{P^N_\pi\in L^N_{\mathrm{PR,SYM}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)-\epsilon_N, \tag{D.2}$$

where $L_{\mathrm{EX}}|_N$ denotes the set of N-DM randomized policies that are the restrictions of policies in L_{EX} to the N first components. By Lemma 1, because L^N_{CO} is convex, without losing global optimality, we can optimize over $L^N_{\mathrm{CO}}\cap L^N_{\mathrm{EX}}$. Let $\epsilon>0$. Consider $P^{*,N}_{\pi,\epsilon}\in L^N_{\mathrm{CO}}\cap L^N_{\mathrm{EX}}$ such that

$$\inf_{P^N_\pi\in L^N_{\mathrm{CO}}\cap L^N_{\mathrm{EX}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \geq \int P^{*,N}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)-\epsilon. \tag{D.3}$$

Following from the proof of Lemma 2, using randomized policies $P^{*,N}_{\pi,\epsilon}\in L^N_{\mathrm{CO}}\cap L^N_{\mathrm{EX}}$, by considering the indexes as a sequence of i.i.d. random variables with uniform distribution on the set $\{1,\dots,N\}$, we can construct an infinitely exchangeable policy $P^{*,\infty}_{\pi,\epsilon}$, where the restriction of an infinitely exchangeable policy to the N first components, $P^{*,\infty}_{\pi,N,\epsilon}\in L^N_{\mathrm{CO}}\cap L_{\mathrm{EX}}|_N$, satisfies

$$\int P^{*,\infty}_{\pi,N,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)\int P^{*,N}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)+\epsilon_N. \tag{D.4}$$

Hence, (D.3) and (D.4) imply that

$$\inf_{P^N_\pi\in L^N_{\mathrm{CO}}\cap L^N_{\mathrm{EX}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)$$

$$\geq \inf_{P^N_\pi\in L^N_{\mathrm{CO}}\cap L_{\mathrm{EX}}|_N}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)-\epsilon-\epsilon_N.$$

Because ϵ is arbitrary, this implies (D.1). By Theorem 1, without losing optimality, we can optimize over $L^N_{\mathrm{CO,SYM}}$. Equality (D.2) is true because $L^N_{\mathrm{CO,SYM}}$ is convex with extreme points in $L^N_{\mathrm{PR,SYM}}$, and the map $\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0): L^N_{\mathrm{CO,SYM}}\to\mathbb{R}$ is linear.

Now, we show that (29) holds. We have

$$\inf_{P^N_\pi\in L^N_{\mathrm{D}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) = \inf_{P^N_\pi\in L^N_{\mathrm{PR}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \tag{D.5}$$

$$\geq \inf_{P^N_\pi\in L^N_{\mathrm{PR,SYM}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)-\epsilon_N, \tag{D.6}$$

where (D.5) follows from Blackwell's [13] irrelevant information theorem because L^N_{CO} is convex with extreme points in L^N_{PR}, and because the map $\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0): L^N_{\mathrm{CO}}\to\mathbb{R}$ is linear, and hence, without losing optimality, we can optimize over L^N_{CO}. Inequality (D.6) follows from (28), and this completes the proof of part i.

To prove part ii, let $P^*_\pi\in L_{\mathrm{PR,SYM}}$ be an optimal policy for $(\mathcal{P}_\infty)$, and $P^*_{\pi,N}$ is the restriction of P^*_π to the first N components. Define, for all $N\in\mathbb{N}$,

$$a_N := \int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0),$$

$$b_N := \inf_{P^N_\pi\in L^N_{\mathrm{PR,SYM}}}\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0).$$

Following from Step 4 of the proof of Theorem 2, because the cost function is bounded, we have

$$\limsup_{N\to\infty} \int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) = \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\mathrm{PR,SYM}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0). \tag{D.7}$$

Hence, $\limsup_{N\to\infty} a_N = \limsup_{N\to\infty} b_N$. Following from Step 4 of the proof of Theorem 2 and symmetry, $\lim_{N\to\infty} a_N = a < \infty$, and also there exists a subsequence such that $\lim_{k\to\infty} b_{N_k} = a < \infty$. On the other hand, because $a_N \geq b_N$ for all $N \in \mathbb{N}$, we can find $\tilde{\epsilon}_N \geq 0$ such that $a_N = b_N + \tilde{\epsilon}_N$. Taking limit as k goes to infinity from both sides, we have $a = \lim_{k\to\infty}(b_{N_k} + \epsilon_{N_k}) = a + \lim_{k\to\infty} \epsilon_{N_k}$. Hence, $\lim_{k\to\infty} \epsilon_{N_k} = 0$ because $\tilde{\epsilon}_N \geq 0$. This implies that there exists $\bar{\epsilon}_N \geq 0$ where $\bar{\epsilon}_N \to 0$ as N goes to infinity such that

$$\int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) \leq \inf_{P^N_\pi \in L^N_{\mathrm{D}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) + \epsilon_N + \bar{\epsilon}_N, \tag{D.8}$$

where (D.8) follows from (29). This completes the proof of part ii. □

References

[1] Aldous DJ, Ibragimov IA, Jacod J (1985) *Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983*, vol. 1117 (Springer, Berlin).
[2] Aliprantis CD, Border KC (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. (Springer, Berlin).
[3] Arabneydi J, Mahajan A (2015) Team-optimal solution of finite number of mean-field coupled LQG subsystems. *Proc. 54th IEEE Conf. Decision Control* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 5308–5313.
[4] Arapostathis A, Biswas A, Carroll J (2017) On solutions of mean field games with ergodic cost. *J. Math. Pures Appliquées* 107(2):205–251.
[5] Arrow KJ, Radner R (1979) Allocation of resources in large teams. *Econometrica* 47(2):361–385.
[6] Banica T, Curran S, Speicher R (2012) De Finetti theorems for easy quantum groups. *Ann. Probab.* 40(1):401–435.
[7] Bardi M, Fischer M (2019) On non-uniqueness and uniqueness of solutions in finite-horizon mean field games. *ESAIM Control Optim. Calculus Variations* 25:44.
[8] Bardi M, Priuli FS (2014) Linear-quadratic *N*-person and mean-field games with ergodic cost. *SIAM J. Control Optim.* 52(5):3022–3052.
[9] Bayraktar E, Zhang X (2020) On non-uniqueness in mean field games. *Proc. Amer. Math. Soc.* 148(9):4091–4106.
[10] Beckmann MJ (1958) Decision and team problems in airline reservations. *Econometrica* 26(1):134–145.
[11] Beneš VE (1971) Existence of optimal stochastic control laws. *SIAM J. Control* 9(3):446–472.
[12] Bertsekas DP, Shreve S (1978) *Stochastic Optimal Control: The Discrete Time Case* (Academic Press, New York).
[13] Blackwell D (1964) Memoryless strategies in finite-stage dynamic programming. *Ann. Math. Statist.* 35:863–865.
[14] Borkar V (2000) Average cost dynamic programming equations for controlled Markov chains with partial observations. *SIAM J. Control Optim.* 39(3):673–681.
[15] Borkar V (2007) Dynamic programming for ergodic control of Markov chains under partial observations: A correction. *SIAM J. Control Optim.* 45(6):2299–2304.
[16] Borkar VS (1993) White-noise representations in stochastic realization theory. *SIAM J. Control Optim.* 31:1093–1102.
[17] Brandao FGSL, Harrow AW (2017) Quantum de Finetti theorems under local measurements with applications. *Comm. Math. Phys.* 353(2): 469–506.
[18] Brunner N, Cavalcanti D, Pironio S, Scarani V, Wehner S (2014) Bell nonlocality. *Rev. Modern Phys.* 86(2):419–478.
[19] Caines P, Huang M, Malhamé R (2017) Mean field games. Başar T, Zaccour G, eds. *Handbook of Dynamic Game Theory* (Springer, Cham), 345–372.
[20] Campi L, Fischer M (2022) Correlated equilibria and mean field games: A simple model. *Math. Oper. Res.*, ePub ahead of print, February 10, https://doi.org/10.1287/moor.2021.1206.
[21] Cardaliaguet P (2011) Notes on mean field games. (from P.-L. Lions' lectures at College de France). Lecture notes, April–May 2010, Tor Vergata, Rome.
[22] Cardaliaguet P, Rainer C (2020) An example of multiple mean field limits in ergodic differential games. *Nonlinear Differential Equations Appl.* 27:25.
[23] Cardaliaguet P, Delarue F, Lasry J, Lions P (2019) *The Master Equation and the Convergence Problem in Mean Field Games.* Annals of Mathematics Studies, vol. 201 (Princeton University Press, Princeton, NJ).
[24] Carmona R, Delarue F (2018) *Probabilistic Theory of Mean Field Games with Applications*, 2 vols. (Springer, Cham, Switzerland).
[25] Carmona R, Delarue F, Lacker D (2016) Mean field games with common noise. *Ann. Probab.* 44(6):3740–3803.
[26] Caves CM, Fuchs CA, Schack R (2002) Unknown quantum states: The quantum de Finetti representation. *J. Math. Phys.* 43(9):4537–4559.
[27] Cecchin A (2021) Finite state *N*-agent and mean field control problems. *ESAIM Control Optim. Calculus Variations* 27:31.
[28] Cecchin A, Pra OD, Fischer M, Pelino G (2019) On the convergence problem in mean field games: A two state model without uniqueness. *SIAM J. Control Optim.* 57(4):2443–2466.
[29] Charalambous CD (2016) Decentralized optimality conditions of stochastic differential decision problems via Girsanov's measure transformation. *Math. Control Signals Systems* 28(3):1–55.
[30] Christandl M, Toner B (2009) Finite de Finetti theorem for conditional probability distributions describing physical theories. *J. Math. Phys.* 50(4):042104.
[31] Davison E, Rau N, Palmay F (1973) The optimal decentralized control of a power system consisting of a number of interconnected synchronous machines. *Internat. J. Control* 18(6):1313–1328.
[32] Delarue F, Tchuendom R (2020) Selection of equilibria in a linear quadratic mean-field game. *Stochastic Processes Their Appl.* 130(2):1000–1040.
[33] Diaconis P, Freedman D (1980) Finite exchangeable sequences. *Ann. Probab.* 8(4):745–764.
[34] Filippov A (1962) On certain questions in the theory of optimal control. *J. Soc. Indust. Appl. Math., Ser. A. Control* 1(1):76–84.

[35] Fischer M (2017) On the connection between symmetric *N*-player games and mean field games. *Ann. Appl. Probab.* 27(2):757–810.
[36] Girsanov IV (1960) On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory Probab. Appl.* 5(3):285–301.
[37] Gupta A, Yüksel S, Başar T, Langbort C (2015) On the existence of optimal policies for a class of static and sequential dynamic teams. *SIAM J. Control Optim.* 53(3):1681–1712.
[38] Hajek B, Livesay M (2019) On non-unique solutions in mean field games. *Proc. 58th IEEE Conf. Decision Control* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 1219–1224.
[39] Hernández-Lerma O, Lasserre JB (1996) *Discrete-Time Markov Control Processes: Basic Optimality Criteria* (Springer, New York).
[40] Hespanha J, Naghshtabrizi P, Xu Y (2007) A survey of recent results in networked control systems. *Proc. IEEE* 95(1):138–162.
[41] Hewitt E, Savage LJ (1955) Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* 80(2):470–501.
[42] Ho Y (1980) Team decision theory and information structures. *Proc. IEEE* 68(6):644–654.
[43] Ho YC, Chu KC (1972) Team decision theory and information structures in optimal control problems—Part I. *IEEE Trans. Automatic Control* 17(1):15–22.
[44] Huang M, Nguyen SL (2016) Linear-quadratic mean field teams with a major agent. *Proc. 55th IEEE Conf. Decision Control* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 6958–6963.
[45] Huang M, Caines PE, Malhamé RP (2006) Large population stochastic dynamic games: Closed-loop Mckean–Vlasov systems and the Nash certainty equivalence principle. *Comm. Inform. Systems* 6(3):221–251.
[46] Huang M, Caines PE, Malhamé RP (2007) Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ-Nash equilibria. *IEEE Trans. Automatic Control* 52(9):1560–1571.
[47] Huang M, Caines PE, Malhamé RP (2012) Social optima in mean field LQG control: Centralized and decentralized strategies. *IEEE Trans. Automatic Control* 57(7):1736–1751.
[48] Jovanovic B, Rosenthal RW (1988) Anonymous sequential games. *J. Math. Econom.* 17(1):77–87.
[49] Kallenberg O (1973) Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 27(1):23–36.
[50] Kallenberg O (2006) *Probabilistic Symmetries and Invariance Principles* (Springer, New York).
[51] Kingman JFC (1978) Uses of exchangeability. *Ann. Probab.* 6(2):183–197.
[52] Krainak JC, Speyer JL, Marcus SI (1982) Static team problems—Part I: Sufficient conditions and the exponential cost criterion. *IEEE Trans. Automatic Control* 27:839–848.
[53] Lacker D (2015) Mean field games via controlled martingale problems: Existence of Markovian equilibria. *Stochastic Processes Their Appl.* 125(7):2856–2894.
[54] Lacker D (2016) A general characterization of the mean field limit for stochastic differential games. *Probab. Theory Related Fields* 165(3–4): 581–648.
[55] Lacker D (2017) Limit theory for controlled Mckean–Vlasov dynamics. *SIAM J. Control Optim.* 55(3):1641–1672.
[56] Lacker D (2020) On the convergence of closed-loop Nash equilibria to the mean field game limit. *Ann. Appl. Probab.* 30(4):1693–1761.
[57] Lasry JM, Lions PL (2007) Mean field games. *Japanese J. Math.* 2:229–260.
[58] Light B, Weintraub GY (2022) Mean field equilibrium: Uniqueness, existence, and comparative statics. *Oper. Res.* 70(1):585–605.
[59] Mahajan A, Martins NC, Yüksel S (2013) Static LQG teams with countably infinite players. *Proc. 52nd IEEE Conf. Decision Control* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 6765–6770.
[60] Mahajan A, Martins N, Rotkowitz M, Yüksel S (2012) Information structures in optimal decentralized control. *Proc. 51st IEEE Conf. Decision Control* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 1291–1306.
[61] Marschak J (1955) Elements for a theory of teams. *Management Sci.* 1(2):127–137.
[62] Mas-Colell A (1984) On a theorem of Schmeidler. *J. Math. Econom.* 13(3):201–206.
[63] McGuire CB (1961) Some team models of a sales organization. *Management Sci.* 7(2):101–130.
[64] Popescu S (2014) Nonlocality beyond quantum mechanics. *Nature Phys.* 10(4):264–270.
[65] Radner R (1962) Team decision problems. *Ann. Math. Statist.* 33(3):857–881.
[66] Renner R (2007) Symmetry of large physical systems implies independence of subsystems. *Nature Phys.* 3(9):645–649.
[67] Saldi N (2019) A topology for team policies and existence of optimal team policies in stochastic team theory. *IEEE Trans. Automatic Control* 65(1):310–317.
[68] Sandell N, Varaiya P, Athans M, Safonov M (1978) Survey of decentralized control methods for large scale systems. *IEEE Trans. Automatic Control* 23(2):108–128.
[69] Sanjari S, Yüksel S (2021a) Optimal policies for convex symmetric stochastic dynamic teams and their mean-field limit. *SIAM J. Control Optim.* 59(2):777–804.
[70] Sanjari S, Yüksel S (2021b) Optimal solutions to infinite-player stochastic teams and mean-field teams. *IEEE Trans. Automatic Control* 66(3):1071–1086.
[71] Sanjari S, Saldi N, Yüksel S (2020) Optimality of independently randomized symmetric policies for exchangeable stochastic teams with infinitely many decision makers. Preprint, submitted August 26, https://arxiv.org/abs/2008.11570.
[72] Schäl M (1975) Conditions for optimality in dynamic programming and for the limit of *n*-stage optimal policies to be optimal. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32:179–296.
[73] Schmeidler D (1973) Equilibrium points of nonatomic games. *J. Statist. Phys.* 7(4):295–300.
[74] Serfozo R (1982) Convergence of Lebesgue integrals with varying measures. *Sankhyā: Indian J. Statist. Ser. A.* 44(3):380–402.
[75] Tsitsiklis JN (1988) Decentralized detection by a large number of sensors. *Math. Control Signals Systems* 1(2):167–182.
[76] Wang BC, Zhang JF (2017) Social optima in mean field linear-quadratic-Gaussian models with Markov jump parameters. *SIAM J. Control Optim.* 55(1):429–456.
[77] Witsenhausen H (1968) A counterexample in stochastic optimal control. *SIAM J. Control Optim.* 6:131–147.
[78] Witsenhausen H (1988) Equivalent stochastic control problems. *Math. Control Signals Systems* 1(1):3–11.

[79] Witsenhausen HS (1975) The intrinsic model for discrete stochastic control: Some open problems. Bensoussan A, Lions JL, eds. *Control Theory, Numerical Methods and Computer Systems Modelling.* Lecture Notes in Economics and Mathematical Systems, vol. 107 (Springer, Berlin), 322–335.
[80] Young L (1937) Generalized curves and the existence of an attained absolute minimum in the calculus of variations. *Comptes Rendus de la Societe des Sci. et des Lettres de Varsovie* 30:212–234.
[81] Yu X, Zhang Y, Zhou Z (2021) Teamwise mean field competitions. *Appl. Math. Optim.* 84:903–942.
[82] Yüksel S (2017) On stochastic stability of a class of non-Markovian processes and applications in quantization. *SIAM J. Control Optim.* 55(2):1241–1260.
[83] Yüksel S (2020) A universal dynamic program and refined existence results for decentralized stochastic control. *SIAM J. Control Optim.* 58(5):2711–2739.
[84] Yüksel S, Başar T (2013) *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints* (Springer, New York).
[85] Yüksel S, Saldi N (2017) Convex analysis in decentralized stochastic control, strategic measures and optimal solutions. *SIAM J. Control Optim.* 55(1):1–28.