Research Statement

Satoshi Fukuda

October 30, 2022

My research fields are microeconomic theory, game theory and information economics. In particular, I am interested in mechanism design, dynamic games, and epistemic game theory. Broadly, my research agenda is to model and elucidate the role that information plays in strategic situations.

Below, I will summarize my research projects based on the research fields: information economics and dynamic games (Section 1) and epidemic game theory (Section 2).

1 Information Economics and Dynamic Games

The first strand of my main research topics is information economics/mechanism design and dynamic games.

1.1 Information Economics

In the field of information economics, I am especially interested in studying how a group of individuals communicate or exchange information to make a better collective decision without resorting to side-payments (e.g., supranational organizations such as a monetary union, political coalitions at the national level, and divisions within a firm).

From Equals to Despots

From Equals to Despots: The Dynamics of Repeated Decision Making in Partnerships with Private Information (with Vinicius Carrasco and William Fuchs, Journal of Economic Theory 2019) considers the optimal dynamic renegotiation-proof mechanism among a group of privately informed agents who repeatedly take a joint common action but who are unable to resort to side-payments. In such a situation, the players may be tempted to exaggerate their preferred actions in order to manipulate the group action. The paper provides a general

framework which accommodates as special cases committee decision and collective insurance problems. While the first-best values can never be attained in an incentive compatible way, the cost of incentives approximately disappears as the agents become patient. In the optimal mechanism, an agent who has a "strong" preference shock can influence a current joint action at the cost of forgoing continuation utilities. Our main result is that the inter-temporal trade-off in the optimal mechanism necessarily leads to a dictatorial mechanism: in the long run, an optimal action takes care only of one player's preferences.

While the above paper studies non-monetary incentives in a dynamic environment, *Who to Listen to?: A Model of Endogenous Delegation* (with William Fuchs and Mahyar Sefidgaran, under review) studies non-monetary incentives in a static environment where it is harder to elicit information from agents.

Who to Listen to?: A Model of Endogenous Delegation

The paper studies a situation in which two players take a joint action without resorting to side-payments. Each player has her own preferred joint action, which is her private information. We study how the range of private information, which is construed as the ex-ante notion of conflict, affects the optimal mechanism.

Examples are abundant. Within a firm, for example, suppose that two departments, sales and engineering, are planning to make a new product design. In a political sphere, consider a coalitional government consisting of two parties.

Our results are as follows. First, we show that an optimal mechanism is deterministic. While introducing randomness may alleviate incentives, it is not needed in our environment. Second, we characterize the optimal mechanism depending on size and location of the support of each player's private types (their preferred action). When there is an excessive amount of conflict, it is too costly to elicit players' information, which leads to an optimal constant allocation. Delegation arises endogenously when there is conflict and asymmetry in the amount of private information. The player with more private information can dictate the allocation with some bounds. In contrast, an overlap of private information leads to information sharing. In this case, committing to sometimes taking ex-post inefficient actions is optimal. The welfare relative to the first-best is non-monotone in the degree of conflict: as the degree of conflict decreases, the relative welfare is at first decreasing and then increasing.

In the field of dynamic games, I am also interested in the role of inter-temporal tradeoffs in applied contexts, in addition to how players communicate or exchange information.

1.2 Dynamic Games: Bargaining, Negotiations and Communication

Negotiations with Limited Specifiability

Negotiations with Limited Specifiability (with Yuichiro Kamada, American Economic Journal: Microeconomics 2022) studies a bargaining problem, and examines how its bargaining protocol affects the set of possible outcomes. The protocol is characterized by three components: (i) when the parties can speak (a proposer rule), (ii) what they can say (a specification rule), and (iii) how they conclude their bargaining (a termination rule).

Especially, the novelty in the paper is to study the role of specification rules. For example, consider negotiations among different countries, say the Conference of the Parties (COP) meetings for climate change. Under the new framework adopted for the 2015 Paris Agreement, each country was able to report their target emission level, while they were not able to specify other countries' emission levels.

While one can study the set of equilibrium bargaining outcomes for each fixed bargaining game, this paper makes it possible to analyze how the bargaining outcomes would change if some bargaining protocols are changed. We show that a bargaining game with alternating announcements leads to a weakly smaller set of outcomes than the corresponding bargaining game with simultaneous proposals. In particular, the outcome is unique when there is a "common interest" alternative. If a specification rule is such that each player may not be able to fully specify a feasible alternative, the set of equilibrium outcomes is larger than the corresponding equilibrium bargaining outcomes under which full specification is possible.

Unprecedented

Unprecedented (with Yuichiro Kamada, preparing for submission) studies a dynamic game in which each player can take a new action only if either she privately learns it or the opponent takes it. For example, a firm may invent a promotion strategy to effectively attract customers, and once this firm employs such a strategy, it may be used by any other firms (think of mileage-based frequent-flyer programs presumably first launched in 1981 by American Airlines).

The new action profile is a Nash equilibrium, and is Pareto dominated by the default action profile. Notable examples are prisoner's dilemma and pure coordination games.

Under the assumptions that taking the new action is an irreversible choice and moves are asynchronous, we show that, when probability of private learning is low and players are patient, there is a unique perfect Bayesian equilibrium. In the unique equilibrium, the new action is never taken, i.e., the new action remains unprecedented. This is the case even

though, after many periods, it is almost common knowledge among the players that they have learned the new action.

1.3 Dynamic Games: Economic Epidemiology

Epidemics with Behavior

In "Epidemics with Behavior" (with Christoph Carnehl and Nead Kos; Conditionally Accepted at Journal of Economic Theory), We study social distancing in an epidemiological model. Distancing reduces the individual's probability of getting infected but comes at a cost. Equilibrium distancing flattens the curve and decreases the final size of the epidemic. We examine the effects of distancing on the outset, the peak, and the final size of the epidemic.

First, the prevalence increases beyond the initial value only if the transmission rate is in the intermediate region. If the transmission rate is too high, individuals distance with such fervor that the prevalence never rises above the initial seed of infection. This finding stands in stark contrast with the predictions offered by the SIR model without distancing where the infection spreads if the transmission rate is high enough.

Second, the peak of the epidemic is non-monotonic in the transmission rate. A reduction in the transmission rate can increase the peak. However, a decrease in the cost of distancing always flattens the curve. Third, both a reduction in the transmission rate as well as a reduction in the cost of distancing decrease the final size of the epidemic.

These two comparative statics lend themselves to two interpretations. Firstly, a disease with a higher transmission rate can lead to a lower peak prevalence. Secondly, a policy that decreases the transmission rate could lead to a higher peak prevalence. In addition, the fact that peak prevalence is monotonic in the cost of distancing and non-monotonic in the transmission rate has important implications on how interventions should be modeled. Namely, public policies that decrease the transmission rate can lead to unintended negative consequences in the short run but not in the long run. Therefore, it is important to distinguish between interventions that affect the transmission rate and interventions that affect contact rates.

Time-varying Cost of Distancing: Distancing Fatigue, Holidays and Lockdowns

In "Time-varying Cost of Distancing: Distancing Fatigue, Holidays and Lockdowns" (with Christoph Carnehl and Nead Kos, a preliminary draft available), We study an SIR model with endogenous behavior and a time-varying cost of distancing. We show that a steep increase in distancing cost is necessary for a second wave of an epidemic to arise. As a special case of the model with changing cost, we study distancing fatigue—the distancing

cost increases in past distancing—and show that it cannot generate a second wave. Moreover, we characterize the change in the distancing cost necessary for the slope of prevalence to change its sign. This characterization informs policymakers: (i) of the required strictness of mitigation policies to cease the increase of prevalence, (ii) when and how policies can be lifted to avoid a second wave, and (iii) whether public holidays are likely to generate another wave of the epidemic. Finally, we illustrate the implementation of desirable time-varying transmission rates through time-varying distancing cost with endogenous equilibrium distancing.

2 Epistemic Game Theory

2.1 Epistemic Game Theory: Reasoning about Belief, Knowledge, and Unawareness: Part 1, Three Representative Papers

The second strand of my main research topics is to formally understand players' belief, knowledge and rationality in a strategic context. Especially, I am interested in studying "boundedly-rational" agents who lack their logical reasoning or introspective abilities. This section discusses three representative papers of mine in this strand of research agenda.

First, *The Existence of Universal Qualitative Belief Spaces* (R&R at Journal of Economic Theory) constructs a canonical representation of players' interactive beliefs about unknown external values such as the payoffs and strategies in a game, irrespective of nature of beliefs—probabilistic (countably-/finitely-/non-additive) or qualitative (qualitative belief or knowledge). That is, the canonical space incorporates all possible ways in which players' interactive beliefs (players' beliefs, players' beliefs about their beliefs, and so forth) are described. Each state of the canonical model encodes players' interactive beliefs at that state within itself in a coherent manner.

While the first paper is a foundational work for analyzing players' interactive beliefs regardless of their underlying properties of beliefs, the second paper, *Formalizing Common Belief with No Underlying Assumption on Individual Beliefs* (Games and Economic Behavior, 2020), studies consequences of strategic reasoning made by not-necessarily perfectly-logical reasoners like humans. To the best of my knowledge, this is the first paper systematically studying consequences of players' common belief in rationality on a solution concept of game theory referred to as an iterated deletion of strictly dominated actions when players are not perfectly logical reasoners. First, if the players in a game are not-necessarily-logical reasoners, then their actions may not necessarily survive an iterated deletion of strictly

dominated actions even if they are rational, they mutually believe their rationality, they mutually believe that they mutually believe their rationality, and so forth *ad infinitum*. Second, the paper proposes the most permissive notion of common belief such that if the players in a game commonly believe their rationality then their resulting actions survive any iterated deletion of strictly dominated actions, irrespective of properties of underlying individual players' beliefs.

Third, when it comes to strategic reasoning, we the outside analysts implicitly assume that the players of a game are "certain" of the structure of the game. While informal arguments exist, to the best of my knowledge, there has been no formalization of such statement. *Are the Players in an Interactive Belief Model Meta-certain of the Model Itself?* (Extended abstract at TARK 2021 Proceedings, preparing for journal submission) formalizes the sense in which the players are "certain" of the structure of a model itself, which has been an implicit and informal assumption in game theory.

2.2 Epistemic Game Theory: Reasoning about Belief, Knowledge, and Unawareness: Part 2

2.2.1 Representations of Probabilistic Beliefs

In a strategic situation in which a player reasons about probabilistic beliefs of the opponents, I am interested in incorporating non-standard notions of probabilistic beliefs. First, *On the Consistency among Prior, Posteriors, and Information Sets* (an extended abstract is available at Electronic Proceedings in Theoretical Computer Science) provides a general framework for capturing both (not-necessarily-countably-additive) probabilistic beliefs and knowledge, extending the standard partitional model of knowledge and countably-additive beliefs. To that end, the paper studies implications of the consistency conditions among prior, posteriors, and information sets on introspective properties of qualitative belief induced from information sets. The main benchmark result characterizes the Bayes law in terms of an agent's introspective abilities: it reformulates the consistency conditions as: (i) the information sets, without any assumption, almost surely form a partition; and (ii) the posterior at a state is equal to the Bayes conditional probability given the corresponding information set. By posing the consistency conditions, one can develop a standard partitional model of knowledge and belief in epistemic game theory without assuming that each partition (information set) has a positive probability. For example, the paper generalizes the famous Agreement and No-Trade theorems to an arbitrary measurable space with the consistency conditions. Next, the paper studies the implications of the consistency conditions. First, each posterior is uniquely determined. Second, qualitative belief reduces to

fully introspective knowledge in a "standard" environment. Thus, a care must be taken when one studies non-veridical belief or non-introspective knowledge. Third, an information partition compatible with the consistency conditions is uniquely determined by the posteriors. Fourth, qualitative and probability-one beliefs satisfy truth axiom almost surely. The paper also sheds light on how the additivity of the posteriors yields negative introspective properties of beliefs.

The next paper concerns about how to conveniently represent players' higher-order beliefs when their beliefs are not necessarily countably additive. In the literature, a p-belief operator is a convenient tool in representing agents' higher-order beliefs. It maps an event E to the event that an agent believes E with probability at least p. By iterating agents' p-belief operators, the analysts can unhold one's beliefs about another's without explicitly constructing beliefs over the space of beliefs. *On p-Belief-Operator Representations of Non-Additive Beliefs* (Preliminary draft) first provides the conditions under which an agent's p-belief operators induce her underlying beliefs at each state of the world, i.e., her type mapping, without any underlying logical assumption on beliefs. Building on this benchmark result, my main objective is to show that p-belief operators alone can be a primitive of an interactive belief model for a wide variety of non-additive beliefs. The representations include Choquet, Dempster-Shafer, and possibility beliefs.

2.2.2 Canonical Representation of Beliefs

As one of my representative papers is on the construction of a canonical representation of players' interactive beliefs, I have been extensively working on constructing a canonical belief representation. First, *A Qualitative Type Space Approach to Hierarchies of Beliefs, Preferences, and Expectations* (Preliminary draft) generalizes a notion of a type. Usually, a type is a probability distribution over the types of the opponent players, and thereby a type induces higher-order beliefs. The paper extends a notion of a type to qualitative beliefs, expectations, and preferences. Take the notion of qualitative beliefs for instance. The paper defines a notion of a qualitative belief type which captures players' interactive qualitative beliefs. The qualitative belief type representation connects the standard type space approach and the possibility correspondence (information set) approach. The paper characterizes various logical and introspective properties of players' qualitative beliefs. Mathematically, this qualitative type approach can also accommodate the standard probabilistic type by considering a collection of qualitative types for each probability. The main result of this paper is to construct a universal hierarchical type space, where types can dictate players' probabilistic/qualitative beliefs, expectations, or preferences.

Second, *Topology-free Constructions of a Universal Type Space as Coherent Belief Hier-*

archies constructs a universal type space on an arbitrary measurable space of nature states as the set of coherent belief hierarchies, proposing the right notion of coherent belief hierarchies. Since any type space induces belief hierarchies of countable depths, coherency in this paper requires that a belief hierarchy (consisting of all finite levels of beliefs) extend to any subsequent countable levels in a way such that all countable levels of beliefs do not conflict with one another. The paper shows that the space of such coherent belief hierarchies is a universal type space without any topological assumption on nature states. Such universal type space coincides exactly with the topology-free universal type space constructed as the set of belief hierarchies that are induced by some type of some type space. Hence, this paper shows that, under the coherency condition that all countable levels of beliefs do not conflict with one another, the previous approaches yield the same universal space in the most general measurable environment without any topological assumption. Moreover, the need for keeping track of all countable levels of beliefs in constructing the universal type space without a topological assumption has a game-theoretic counterpart: the need for transfinite levels of reasoning (e.g., eliminations of strictly dominated actions) in solving infinite games with general measurable action spaces employing rationalizability solution concepts.

2.2.3 Reasoning about Unawareness

The study of agents who lack introspective abilities naturally led to unawareness. There are two ways to approach unawareness in the existing literature. One is to define unawareness as a lack of knowledge: an agent is unaware of a statement if she does not know it and she does not know that she does not know it. The other is to define unawareness as a lack of "concept."

Unawareness without AU Introspection (Journal of Mathematical Economics, 2021) provides a general model that nests and allows for comparing both approaches in a unified way. It studies: (i) when two approaches lead to non-trivial forms of unawareness; (ii) when two approaches coincide with each other; (iii) when an agents is aware of being unaware of "something;" and (iv) getting more information may cause an agent to be less aware (i.e., when an agent is not fully introspective, the value of information may be negative).

Strategic Games with Possibility Correspondence Models of Unawareness (Preliminary Draft) axiomatizes a possibility correspondence model of unawareness on a generalized state space by underlying properties of beliefs, and provides an epistemic characterization of iterated elimination of strictly dominated actions (IESDA) in a game with unawareness as an implication of common belief in rationality. First, the paper axiomatizes a wide variety of unawareness structures that respect given desirable properties of beliefs. Specifically, I fully characterize properties of a possibility correspondence that yields the corresponding

properties of the induced belief operator. Conversely, I analyze conditions on a given belief operator which generate a well-defined possibility correspondence, which, in turn, induces the original belief operator. Second, irrespective of properties of beliefs, if players commonly believe their rationality, then their resulting actions survive IESDA even with the presence of unawareness. However, unawareness may increase the set of actions that are consistent with common belief in rationality. I also identify a property of unawareness under which a player may be unaware of her own rationality even if she is rational.

2.2.4 Interdisciplinary Works on Representations of Knowledge and Belief

Finally, modeling decision-makers' beliefs, knowledge, and unawareness pertains to such various fields as computer science, logic, philosophy, and psychology as well as economics and game theory. First, in such fields, an agent's knowledge is informally summarized by a collection of sets such as a topology or a σ-algebra (see, for example, a standard textbook on measure and probability theory such as Billingsley (2012) "Probability and Measure." Anniversary Edition. Wiley). *Epistemic Foundations for Set-algebraic Representations of Knowledge* (Journal of Mathematical Economics 2019) formalizes such informal idea, and fully characterizes why the agent's knowledge takes (or does not take) such a set algebra as a σ-algebra or a topology, depending on logical and introspective properties of knowledge and on the underlying structure of the state space.

Second, an agent's knowledge is also represented by a single information set in computer science, economics, logic, and philosophy. *An Information Correspondence Approach to Bridging Knowledge-Belief Representations in Economics and Mathematical Psychology* develops a model of interactive beliefs and knowledge which I call an information correspondence. The information correspondence assigns multiple information sets at each state. This generalization allows one to analyze an agent who fails to believe the conjunction of her own beliefs or a tautology. This generalization also enables one to study qualitative and probabilistic beliefs in a unified manner (the standard single-information-set approach can only represent qualitative beliefs). The model nests a knowledge representation in mathematical psychology known as a surmise function.

Third, *Can the Crowd be Introspective? Modeling Distributed Knowledge from Collective Information through Inference* studies a notion of "distributed knowledge" among a group of agents who possibly have contradictory beliefs with each other. While collective knowledge is at the heart of the market system, the paper focuses on how one can formally define and represent group knowledge, the questions that would be asked by computer scientists and philosophers rather than economists. The paper formalizes distributed knowledge as knowledge logically deduced from agents' collective information, consisting of events that some

agent believes whenever they are true. Roughly, first, a group of agents can be collectively unaware of events—if the group, as a whole, does not know something, the group may not know that the group does not know it. Second, if agents' beliefs are true, monotonic, positively introspective, and conjunctive, then distributed knowledge coincides with knowledge possessed by the least knowledgeable "wise man" who knows everything each agent knows.