

“Group Beliefs”: Conversation Transcript

Benjamin Brooks, Songzi Du, and Claude

September 17, 2026

This document is a raw transcript of the conversation in which the working paper “Group Beliefs” (Brooks and Du) was drafted. Messages are reproduced verbatim and in order. Tool activity (file inspection, LaTeX compilation, and web searches) is omitted; deliveries of files are noted in bracketed italics. The conversation was conducted from Ben’s account, but on the human side both Ben and Songzi took part; the human turns are therefore attributed jointly.

Ben & Songzi

[Attached: seconddraft_9-15-2026_bab.tex, references.bib]

I would like to write a new paper and I want your assistance with it. You have seen a number of my papers and my customary style. The paper is joint with Songzi Du and it is on the topic of group belief invariant BCE, which I discussed with you in another recent chat, while writing a referee report for me on the paper “Robust Predictions with Bounded Information.” The subject is to further explore group belief invariance. The title of the paper is “Group Beliefs”. The first step is to introduce all the standard notation for a Bayesian game in a model section. Please create a tex file for a new paper. For comparison, I have uploaded the tex file for the robust predictions paper, along with the bib file of references.

Claude

The skeleton is below. A few notes on the choices I made:

Model section. It follows the bounded-information paper’s notation exactly — players $i = 1, \dots, N$, finite Θ , base information structure $\underline{I} = (\underline{S}, \underline{\sigma})$, expansions, game structure $G = (A, u)$, strategies, induced outcomes $\phi \in \Delta(A \times \underline{S} \times \Theta)$, and BCE via obedience constraints with the BM (2016) epistemic result. Two additions relative to a straight transcription: (i) I introduce groups J and the x_J/x_{-J} notation up front, since it’s the paper’s namesake; (ii) the section culminates in the definition of group belief invariant BCE, with the qualifier “for all $(\underline{s}_{-J}, \theta)$ with positive probability given $\underline{s}_J$ ” made explicit, followed by a paragraph identifying the singleton and grand-group extremes with belief invariant BCE (Forges 1993, 2006; BM 2016; Liu 2015) and the Bayesian solution. I dropped the feasibility-correspondence material (F_I, F_T) since IGC machinery presumably won’t carry over, but kept $E_T(G)$.

Bibliography. I copied your bib and appended a brooksdu2026bounded working-paper entry for Brooks–Du–Haberman, cited at the definition. Adjust the institution field or swap in an arXiv/SSRN reference as appropriate.

Placeholders. Provisional one-sentence abstract, plus Introduction/Results/Conclusion stubs. The title footnote credits NSF #2215259 with your and Songzi’s affiliations; I omitted the Claude acknowledgment and RA lines pending your call.

One open question worth deciding early: whether the paper should also fix the $N = 2$ coincidence with the belief invariant Bayesian solution in the model section (as in the proposition we discussed for the referee report), or defer all results to Section 3. I left it out of the model for now.

[Files delivered: groupbeliefs.tex, references.bib]

Ben & Songzi

We do not need the base information. Could you compile the pdf?

Claude

Compiles cleanly (I test-compiled with plainnat substituted for ecta and mathabx commented out, since neither is in this container — the delivered file keeps your original preamble, which will work on your machine).

Done — revised tex and compiled PDF are below. The base information structure is gone: I fixed a prior $\mu \in \Delta(\Theta)$, information structures are consistent with μ , strategies are $b_i : S_i \rightarrow \Delta(A_i)$, outcomes live in $\Delta(A \times \Theta)$ with marginal μ , and the BCE obedience constraints lose the s_i index. The bibliography now compiles with the actual ecta.bst (fetched a copy for the container). The Liu (2015) hierarchy characterization was dropped since it was phrased in terms of expansions.

One substantive point you should look at. Without base signals, group belief invariance is the requirement that $a_J \perp \theta$ for every group J , and since independence is preserved under marginalization, the grand-group condition implies all the others. So in this setting group belief invariant BCE collapses to the Bayesian solution — the group structure does no work. I recorded this as an observation at the end of the model section rather than suppressing it, but it suggests either (i) the paper’s interesting content lives elsewhere (e.g., in conditioning on the groups’ signals within the information structure rather than on outcomes, or in per-group information bounds that are positive rather than zero), or (ii) some conditioning variable should replace the base signals in the definition. Happy to restructure the definition once you tell me which direction the paper takes.

Two mechanical notes: mathabx is now loaded via an `\IfFileExists` guard since it’s unused so far and unavailable here — revert to a plain `\usepackage` if you prefer; and the working-paper bib entry brooksdu2026bounded for the bounded-information paper is still appended at the end of references.bib.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

Add a definition of a “group information structure”, which is just an information structure where the “players” are indexed by subsets of $\{1, \dots, N\}$, referred to as groups, and each group has degenerate beliefs about the signals of groups that are subsets of itself.

Claude

The definition was already in the file I delivered last turn (lines 95–109) — it may not have made it into the copy you’re looking at. It’s there in full: signals indexed by groups $J \in \mathcal{J}$, signal space $S = \prod_{J \in \mathcal{J}} S_J$, joint distribution $\tau \in \Delta(S \times \Theta)$ with marginal μ , and the degeneracy condition for nested groups, stated both in conditional-distribution form and via the equivalent measurability formulation with mappings $f_{J'J} : S_J \rightarrow S_{J'}$. Let me add one interpretive sentence in the style of the bounded-information paper and recompile so you have a current PDF.

Updated files are below. The definition sits between the ordinary information structure and the game structure: a group information structure $\mathbb{I} = (T, \tau)$ has signal space $T = \prod_{J \in \mathcal{J}} T_J$ indexed by the set of groups $\mathcal{J}$, joint distribution $\tau \in \Delta(T \times \Theta)$ with marginal μ , and the requirement that for all $J' \subseteq J$, each positive-probability t_J induces a degenerate conditional distribution over $t_{J'}$ — equivalently, $t_{J'} = f_{J'J}(t_J)$ almost surely for some mapping $f_{J'J} : T_J \rightarrow T_{J'}$. I added one sentence of interpretation: the signal of a group determines the signals of each of its subgroups, including the singletons corresponding to its individual members.

Two design points you may want to settle when the results section takes shape. First, as written, T_J is not required to be the join of its members’ signals — a group may know strictly more than the pooled information of its singletons (and of all its proper subgroups). If group signals should be exactly the members’ pooled information, that’s an additional condition, e.g., that t_J be measurable with respect to $(t_{\{i\}})_{i \in J}$. Second, the nested-degeneracy condition automatically delivers consistency of the mappings, $f_{J''J} = f_{J''J'} \circ f_{J'J}$ almost surely for $J'' \subseteq J' \subseteq J$, so nothing further is needed there.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

That is how I wanted you to define it. Please switch to s/S rather than t/T for the signals.

Claude

Done — the group information structure now reads $\mathbb{I} = (S, \tau)$ with group signal space $S = \prod_{J \in \mathcal{J}} S_J$, signals s_J , and mappings $f_{J'J} : S_J \rightarrow S_{J'}$ with $s_{J'} = f_{J'J}(s_J)$ almost surely. No stray t ’s remain, and it compiles cleanly (updated files below).

Note the deliberate overload this creates: $S = \prod_{i=1}^N S_i$ for an ordinary information structure and $S = \prod_{J \in \mathcal{J}} S_J$ for a group information structure, disambiguated by whether the index is a player or a group. That reads naturally — a player i ’s signal is just the signal of the singleton group, modulo identifying s_i with $s_{\{i\}}$ — but if you later want to treat an ordinary information structure as the restriction of a group information structure to singletons, it may be worth saying explicitly that s_i and $s_{\{i\}}$ are identified.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

I see the notational conflict but it is acceptable for now. I actually decided that I would like to use the terminology “type space” instead of “information structure” with t/T instead of s/S. Could you make this change throughout the paper?

Claude

Done — updated files below. “Information structure” is now “type space” throughout, with types t and type sets T : the ordinary type space is $I = (T, \sigma)$ with $T = \prod_{i=1}^N T_i$, the group type space is $\mathbb{I} = (T, \tau)$ with $T = \prod_{J \in \mathcal{J}} T_J$ and mappings $f_{JJ'} : T_J \rightarrow T_{J'}$, strategies are $b_i : T_i \rightarrow \Delta(A_i)$, and the expected-utility and induced-outcome expressions sum over $t \in T$. No occurrences of “information structure” or “signal” remain, and it compiles cleanly to the same 5 pages.

Two things I held fixed pending your preference: the distribution letters (σ for the type space, τ for the group type space, as before), and the structure symbols I and $\mathbb{I}$, which were mnemonic for “information” — if you’d rather have, say, $\mathcal{T} = (T, \sigma)$ and $\mathbb{T} = (T, \tau)$, that’s a quick change.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I would like to have (T, σ) .

Claude

Done — the group type space is now $\mathbb{I} = (T, \sigma)$, with $\sigma \in \Delta(T \times \Theta)$ and the degeneracy conditions stated under σ ; no τ remains anywhere in the file. Both kinds of type space now share the (T, σ) notation, distinguished by whether the index set is players or groups. Updated tex and PDF are below.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Next, I would like you to create a new section titled “The universal group type space”. We will construct the eponymous type space. Start by following the steps in Brandenburger and Dekel (1993) to construct a universal type space, but where the “players” are again subsets of $\{1, \dots, N\}$.

Claude

The new Section 3 is in place (files below), now 7 pages. It follows Brandenburger–Dekel step by step with the player set replaced by $\mathcal{J}$, $M = |\mathcal{J}| = 2^N - 1$:

1. Topological preliminaries ($\Delta(X)$ with the weak* topology, Polish spaces closed under countable products), with a note that this section drops the finiteness restriction on type sets from Section 2.
2. The spaces $X_0 = \Theta$, $X_{n+1} = X_n \times (\Delta(X_n))^{M-1}$, hierarchies $t = (\delta_1, \delta_2, \dots) \in H_0 = \prod_n \Delta(X_n)$, with a parenthetical noting the X_n are group-independent since every group has $M - 1$ others, coordinates indexed by $\mathcal{J} \setminus \{J\}$.
3. Coherency ($\text{marg}_{X_{n-1}} \delta_{n+1} = \delta_n$) and Proposition 1: a homeomorphism $f : H_1 \rightarrow \Delta(\Theta \times H_0^{M-1})$ preserving marginals, with proof deferred to BD’s Proposition 1 and a one-sentence gloss that the substance is Kolmogorov extension.
4. Common knowledge of coherency via $H_{k+1} = \{t \in H_k : f(t)(\Theta \times H_k^{M-1}) = 1\}$, $T^u = \bigcap_k H_k$, and Proposition 2: $g = f|_{T^u}$ is a homeomorphism onto $\Delta(\Theta \times (T^u)^{M-1})$.

The section closes by noting that T^u is just the universal type space for the player set $\mathcal{J}$ — the degeneracy requirement from Section 2 hasn't yet been imposed — and that the universal group type space will be carved out of T^u by imposing that requirement and its common knowledge. That sentence is a placeholder for the next step; cut or rephrase if you have a different route in mind.

I added `brandenburger1993hierarchies` and `mertens1985formulation` to the bib. One modeling note: BD work with a Polish basic space S ; since Θ is finite here, all their regularity hypotheses hold trivially, so the propositions transfer without qualification. If you later want Θ Polish rather than finite (e.g., for the linear-game applications), nothing in the section needs to change except the sentence assigning Θ the discrete topology.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*, *references.bib*]

Ben & Songzi

Now construct the subset of the set T^u in which it is common knowledge that each set of players knows the beliefs of its subsets. This can be done analogously to how you imposed coherence. First define the subset of T^u in which all sets have degenerate beliefs about subsets. Then construct a sequence of decreasing sets of types, in which every type's belief is concentrated on the previous set. This converges to a limit. Finally, assign notation to the subset of the limit which is consistent with the common prior. We will call this the universal group type space.

Claude

The construction is complete and compiles cleanly to 8 pages (files below). Section 3 now proceeds in the three steps you specified:

1. **Degeneracy.** For a type $t \in T^u$ of group J , $g_{J'}(t) \in \Delta(T^u)$ is the marginal of $g(t)$ on the J' -coordinate, and $D_J^1 = \{t \in T^u : g_{J'}(t) \text{ degenerate for every } J' \subsetneq J\}$ — the types of J that know the types, hence the beliefs, of J 's strict subsets.
2. **Common knowledge.** $D_J^{k+1} = \{t \in D_J^k : g(t)(\Theta \times \prod_{J' \neq J} D_{J'}^k) = 1\}$, decreasing in k , with limit $D_J^\infty = \bigcap_k D_J^k$. I included two short verifications: each D_J^k is closed (degenerate measures form a closed set, g and $g_{J'}$ are continuous, then induction), and by continuity from above every $t \in D_J^\infty$ concentrates on $\Theta \times \prod_{J'} D_{J'}^\infty$ — so on $\prod_J D_J^\infty$ it is common knowledge that every group knows its subgroups' beliefs.
3. **Consistency with the prior.** A common prior is $\pi \in \Delta(\Theta \times \prod_J D_J^\infty)$ whose conditional of (θ, t_{-J}) given t_J is $g(t_J)$ for π -a.e. t_J , for every J ; it is consistent with μ if $\text{marg}_\Theta \pi = \mu$. Then $\mathbb{T} = \{t \in \prod_J D_J^\infty : t \in \text{supp } \tilde{\pi} \text{ for some common prior } \pi \text{ consistent with } \mu\}$, christened the universal group type space, with a closing remark that $\mathbb{T} \neq \emptyset$ via the profile where it is common knowledge that $\theta \sim \mu$ and no group knows more.

Two choices you may want to revisit. First, D_J^1 requires degeneracy only for *strict* subsets $J' \subsetneq J$, since $g(t)$ has no J -coordinate — the reflexive case from Section 2 is vacuous here. Second, $\mathbb{T}$ is a set of type *profiles*, not a product of per-group type sets, so it is not literally a group type space in the sense of Section 2 (which required $T = \prod_J T_J$); if you want that reconciliation, one option is to note that any consistent π restricted to the closure of its support defines a (possibly non-product) type space, or to define $\mathbb{T}_J$ as the J -projection. Also worth deciding later whether $\mathbb{T}$ should be the union of supports (as now, following Mertens–Zamir's consistent set) or the closure of that union.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

I would now like you to define projections from group type spaces to the group universal type space. This is the same projection defined in mertens-zamir and brandenburger-dekel. Verify that the projection diagram commutes: Beliefs about others types, and then projecting gives the same distribution over universal types as if one first projects and then applies the homeomorphism.

Claude

The projections are in place and the paper compiles cleanly to 10 pages (files below). The treatment closes Section 3 in three pieces:

Definition. For a group type space $\mathbb{I} = (T, \sigma)$ (with the harmless normalization that every type has positive probability), $\beta_J(t_J) \in \Delta(\Theta \times T_{-J})$ is type t_J 's belief, and the projections $h_J : T_J \rightarrow T^u$ are built by the MZ/BD simultaneous induction: auxiliary maps $\psi_J^n : \Theta \times T_{-J} \rightarrow X_n$ starting from $\psi_J^0(\theta, t_{-J}) = \theta$, with $h_J^{n+1}(t_J) = \beta_J(t_J) \circ (\psi_J^n)^{-1}$ the $(n+1)$ th-order belief and $\psi_J^{n+1} = (\psi_J^n, (h_J^{n+1})_{J' \neq J})$.

Commutation (Proposition 3). $g(h_J(t_J)) = \beta_J(t_J) \circ (\text{id}_\Theta \times h_{-J})^{-1}$: beliefs-then-project equals project-then-homeomorphism. The proof runs through the finite-order projections $q_n : \Theta \times H_0^{M-1} \rightarrow X_n$: the key identity is $q_n \circ (\text{id}_\Theta \times h_{-J}) = \psi_J^n$, which gives $\text{marg}_{X_{n-1}} \nu = h_J^n(t_J)$ for the pushed-forward belief ν ; Proposition 1 gives the same marginals for $f(h_J(t_J))$; and the cylinder sets form a generating π -system, so the two measures coincide. Coherency of $h_J(t_J)$ and membership in T^u (common knowledge of coherency, by induction on H_k using that ν is supported on projection images) are established along the way, so g applies.

Image (Proposition 4). $h_J(T_J) \subseteq D_J^\infty$: the commutation identity plus the degeneracy property of $\mathbb{I}$ gives $g_{J'}(h_J(t_J))$ degenerate at $h_{J'}(f_{J'J}(t_J))$ for $J' \subsetneq J$, hence D_J^1 , and the same induction as for coherency delivers D_J^∞ . Then the pushforward of σ under $(t, \theta) \mapsto (\theta, h(t))$ is a common prior consistent with μ — the conditional given u_J is a mixture over $t_J \in h_J^{-1}(u_J)$ of measures that each equal $g(u_J)$ by commutation, so redundant types cause no trouble — and finite support puts $h(t) \in \mathbb{T}$ for every positive-probability profile.

One notation flag worth a decision at some point: f now does double duty — the BD homeomorphism on H_1 and the Section 2 subgroup maps $f_{J'J}$ — and both appear in the proof of Proposition 4. They're distinguishable by the subscripts, but renaming one (say the Section 2 maps to $\kappa_{J'J}$, or the homeomorphism to β^*) would remove the friction. Similarly h is now the projection while h_J^n carries the order superscript, which is fine, but tell me if you'd rather reserve a different letter.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

Please delete the section called “results”

Claude

Done — the Results section is deleted, with no dangling references to it. The paper now runs Introduction, Model, The universal group type space, Conclusion, and compiles cleanly. Updated files are below.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Now we need a new section on “Group beliefs and equilibrium”. Given a group type space and a game, define an outcome to be a joint distribution over the players’ actions, each group’s type, and the state. Define what it means for a group type space and a profile of strategies of the players to induce an outcome.

Claude

Section 4, “Group beliefs and equilibrium,” is in place (files below). Given $G = (A, u)$ and $\mathbb{I} = (T, \sigma)$: actions are taken by the individual players, with each player i observing the type $t_{\{i\}}$ of their singleton group, so strategies are $b_i : T_{\{i\}} \rightarrow \Delta(A_i)$, identified with $b : \prod_i T_{\{i\}} \rightarrow \Delta(A)$ as usual. An outcome is $\phi \in \Delta(A \times T \times \Theta)$ — actions, each group’s type, and the state — with marginal σ on $T \times \Theta$, and $\mathbb{I}$ and b induce $\phi(a, t, \theta) = \sigma(t, \theta) \prod_i b_i(a_i | t_{\{i\}})$, followed by a one-line check that induced outcomes satisfy the marginal condition.

One modeling choice I made that you should confirm: players act on their *singleton-group* types only. Since the type of any group containing i determines $t_{\{i\}}$ but not conversely, this is the natural reading — the group type space describes what coalitions jointly know, but individuals can only act on what they individually know. The alternative, where a strategy conditions on the types of every group containing i , would credit each player with their groups’ pooled information; if that’s the intended interpretation for some results, the strategy domain changes. Also note “outcome” now has a second meaning relative to Section 2’s $\Delta(A \times \Theta)$ — the marginal of a Section 4 outcome on $A \times \Theta$ is an outcome in the Section 2 sense, which will presumably be the bridge when the two solution concepts get connected; happy to add a sentence making that identification explicit whenever it becomes load-bearing.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Back in Section 2, please add the following construction and definition: A type space (T, σ) induces a group type space in a natural way, where the group’s type is just the vector of types of its members. Prove that this is in fact a group type space.

Claude

The construction is in Section 2, right after the group type space definition (files below, now 11 pages). Given a type space $I = (T, \sigma)$, the induced group type space is $\mathbb{I}(I) = (\hat{T}, \hat{\sigma})$ with $T_J = \prod_{i \in J} T_i$ — which dovetails with the profile notation $t_J = (t_i)_{i \in J}$ already established at the top of the section — and $\hat{\sigma}$ the distribution of $((t_J)_{J \in \mathcal{J}}, \theta)$ under σ . The proposition and proof verify both requirements: $\mathbb{I}(I)$ is a type space with players indexed by groups whose Θ -marginal is μ , and

the degeneracy condition holds with $f_{J'J}$ equal to the coordinate projection $(t_i)_{i \in J} \mapsto (t_i)_{i \in J'}$, since every point in the support of $\hat{\sigma}$ consists of coordinates that are all projections of a single underlying profile t .

The subsequent propositions renumbered automatically (the universal-space results are now Propositions 2–5). One remark for later: the induced group type space is the extreme case in which every group knows exactly the pooled information of its members — no more. Since your general definition permits t_J to encode strictly more than $(t_i)_{i \in J}$, the induced construction presumably ends up characterizing a distinguished subclass (something like “distributed” or “pooling” group type spaces), and it may eventually be worth a remark that in $\mathbb{I}(I)$, the type of a group is measurable with respect to its members’ singleton types — the converse inclusion to the degeneracy condition. Say the word if you’d like that remark added now.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Revise the definition of an outcome. I now want to define an outcome as just a joint distribution over actions and states, and integrate out the types.

Claude

Revised (files below). Section 4 now points back to the Section 2 definition — an outcome is $\phi \in \Delta(A \times \Theta)$ with marginal μ on Θ — and the induced outcome integrates out the types:

$$\phi(a, \theta) = \sum_{t \in T} \sigma(t, \theta) \prod_{i=1}^N b_i(a_i \mid t_{\{i\}}),$$

with the interpretive sentence adjusted accordingly and the marginal check now on Θ .

A pleasant consequence of this revision: outcomes are now a single unified object across the paper, so the sets $E_I(G)$ from Section 2 and the equilibrium outcomes of group type spaces (once equilibrium is defined) will live in the same space $\Delta(A \times \Theta)$ and can be compared directly — presumably where the group belief invariance characterization will land.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Here is a conjecture. For a type space T , let $S(T)$ be the set of distributions over actions and states that are marginals of group belief-invariant BCE. Here is a conjecture: Suppose that T and T' project to the same distribution in the universal group type space. Then $S(T)=S(T')$. Is this conjecture correct?

Claude

Yes, I believe the conjecture is correct, and the machinery already in the paper is exactly what proves it. Here is the argument, with the one genuinely delicate point spelled out.

Setup. Take the BDH definition of group belief invariance relative to a base type space T : a BCE is $\phi \in \Delta(A \times T \times \Theta)$ with marginal σ , obedience conditional on $(t_{\{i\}}, a_i)$, and for every group J ,

$a_J \perp (t_{-J}, \theta)$ conditional on $t_J = (t_i)_{i \in J}$. Let π be the common projection of $\mathbb{I}(T)$ and $\mathbb{I}(T')$, and define the intermediate object $S^*(\pi)$: distributions $\psi \in \Delta(A \times \text{supp } \pi \times \Theta)$ with (u, θ) -marginal π , obedience conditional on $(u_{\{i\}}, a_i)$, and $a_J \perp (u_{-J}, \theta)$ given u_J for every J . The claim follows from $S(T) = S^*(\pi)$, proved in both directions, since S^* depends on T only through π .

Pushdown ($S(T) \subseteq S^*(\pi)$). Push a GBI BCE ϕ forward under $(a, t, \theta) \mapsto (a, h(t), \theta)$. Obedience aggregates: the constraint at $(u_{\{i\}}, a_i)$ is the sum over the fiber $h_{\{i\}}^{-1}(u_{\{i\}})$ of the constraints at $(t_{\{i\}}, a_i)$, each nonnegative. Group belief invariance is where Proposition 3 does the work: conditional on t_J , ϕ 's invariance gives $a_J \perp (u_{-J}, \theta)$ (since u_{-J} is a function of (t_J, t_{-J})), so the joint of (a_J, u_{-J}, θ) given t_J is $P(a_J | t_J) \otimes g(u_J)$ — and by the commutation identity, the second factor is *the same measure* $g(u_J)$ for every t_J in the fiber. A mixture of product measures with a common second marginal is a product, so $a_J \perp (u_{-J}, \theta)$ given u_J .

Lift ($S^*(\pi) \subseteq S(T)$). Given ψ , define $\phi(a | t, \theta) = \psi(a | h(t), \theta)$. Invariance is immediate: ψ 's invariance means $\psi(a_J | u, \theta)$ depends only on $u_J = h_J(t_J)$, a function of t_J , so a_J 's conditional is constant in (t_{-J}, θ) . The delicate point is obedience, because $t_{\{i\}}$ is *finer* than $u_{\{i\}}$ when T has redundant types, and finer conditioning can break averaged incentive constraints. It doesn't here, and the reason is again the commutation identity plus the common-prior property in Proposition 4: conditional on $t_{\{i\}}$, the distribution of $(\theta, u_{-\{i\}})$ is $g(u_{\{i\}})$, which coincides with π 's conditional given $u_{\{i\}}$. Since the lifted recommendation depends on t only through (u, θ) , player i 's joint belief over (a, θ) given $t_{\{i\}}$ is *identical* across the fiber and equals the ψ -belief given $u_{\{i\}}$. So the redundant distinction carries no information the mediator uses, and obedience at $(t_{\{i\}}, a_i)$ is equivalent to obedience at $(u_{\{i\}}, a_i)$. Marginals on $A \times \Theta$ are preserved in both directions, so $S(T) = S^*(\pi) = S(T')$.

Two remarks. First, this is the group analogue of Liu (2015): group hierarchies, not individual ones, are the sufficient statistic for GBI predictions, and the hypothesis genuinely cannot be weakened to individual hierarchies. The XOR example makes this sharp: $\Theta = \{0, 1\}$ uniform, t_1, t_2 i.i.d. uniform with $t_1 \oplus t_2 = \theta$. Every individual hierarchy is common-knowledge-of-uniform, identical to the no-information type space, but the grand group knows θ , so its invariance constraint ($a \perp \theta$ given t) is vacuous and S contains state-matching outcomes that the no-information space excludes. Second, a bookkeeping caveat: the paper's current Section 2 defines GBI without base types (where, as we noted, it degenerates to the Bayesian solution), so writing this up as a theorem requires the with-base-types definition — conditioning on t_J — which is presumably where the group-type-space and outcome machinery of Sections 2–4 has been heading. The proof would cite exactly Propositions 3 and 4; nothing new is needed beyond the fiber-mixing arguments above.

The result also does what you'd want structurally: it licenses defining GBI BCE directly on the universal group type space and treating S as a function on distributions over $\mathbb{T}$. Happy to draft this as a theorem with full proof in the paper whenever you'd like.

Ben & Songzi

Go ahead and draft it as a theorem.

Claude

The theorem is drafted in Section 4 (files below, now 13 pages). The added material, in order:

Definitions. BCE of $(G, \mathbb{I})$: $\phi \in \Delta(A \times T \times \Theta)$ with marginal σ , satisfying obedience conditional

on each player’s singleton type and recommendation (equation 4). Group belief invariance: for every group J , $a_J \perp (t_{-J}, \theta)$ conditional on t_J , stated in the same constant-conditional form as Section 2. Then $S_{\mathbb{I}}(G)$ is the set of $A \times \Theta$ -marginals of GBI BCE, with two connective remarks — elements of $S_{\mathbb{I}}(G)$ are outcomes, and singleton type sets recover the Section 2 concept. The projection distribution $\pi_{\mathbb{I}}$ is named as the common prior from Proposition 5.

Theorem 1. Same projection distribution implies $S_{\mathbb{I}}(G) = S_{\mathbb{I}'}(G)$ for every G , framed as the group analogue of Liu (2015) with a sentence noting the type-space corollary via induced group type spaces (your original conjecture).

Proof. Via the auxiliary set $S^*(G, \pi)$ of GBI-and-obedient distributions on $A \times U \times \Theta$, $U = \text{supp } \bar{\pi}$, showing $S_{\mathbb{I}}(G) = S^*(G, \pi)$; since S^* depends on $\mathbb{I}$ only through π , symmetry finishes it. The proof isolates the two inputs up front — the commutation identity (conditional law of $(\theta, h_{-J}(t_{-J}))$ given t_J is $g(h_J(t_J))$, hence fiber-constant) and the common-prior property of π . Pushdown: obedience aggregates over fibers as a sum of nonnegative constraints; invariance survives because each fiber component is a product with the *common* second factor $g(v_J)$. Lift: $\phi(a, t, \theta) = \sigma(t, \theta)\psi(a \mid h(t), \theta)$; invariance is immediate since ψ ’s conditional depends only on $u_J = h_J(t_J)$; and the finer-conditioning worry for obedience resolves through the identity $\Pr(\theta, u_{-\{i\}} \mid \tau) = g(v)(\theta, u_{-\{i\}})$, making the type- τ constraint a positive multiple $\sigma(\tau) / \bar{\pi}_{\{i\}}(v)$ of the corresponding universal constraint.

Everything compiles with no errors, bad boxes, or dangling references. One presentational thought: the proof is currently inline and runs about a page and a half; if you’d rather match the BDH structure, I can move it to an omitted-proofs appendix and leave a proof sketch in the body. And if you want the sufficiency framing sharpened later, the natural companion result is the converse direction — that the projection distribution is essentially the *coarsest* sufficient statistic — for which the XOR-style example we discussed would be the seed.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Add the xor type space example in the introduction and an explanation of why it has the same belief hierarchies as no information. Add a reference to Ely and Peski (2006) and to Harsanyi (1967) and Myerson’s (2004) commentary.

Claude

The introduction now opens with three paragraphs (files below, 14 pages, all citations resolving):

1. **Framing.** Harsanyi (1967) types encode beliefs about the state and about others’ types, hence full hierarchies (Mertens–Zamir; Brandenburger–Dekel), with Myerson’s (2004) commentary cited as developing that perspective.
2. **The XOR example.** θ uniform on $\{0, 1\}$, t_1 uniform and independent of θ , $t_2 = t_1 \oplus \theta$. The hierarchy argument is spelled out by induction: each type is independent of the state, so first-order beliefs are uniform at every type; therefore each player is certain of the other’s first-order belief regardless of type, so second-order beliefs are type-independent; and so on up the hierarchy — the hierarchies coincide exactly with those of the no-information type space. Yet $\theta = t_1 \oplus t_2$, so the pair jointly knows the state. Attributed as “a version of this

example appears in Ely and Pęski (2006),” with their use of it (individual hierarchies don’t determine interim rationalizable behavior) noted.

3. **The punchline.** What individual hierarchies miss is what groups jointly know, which is the paper’s subject — ending with the [TO BE CONTINUED] marker for the rest of the intro (contributions, preview of the universal construction and Theorem 1, related literature).

I added `ely2006hierarchies` and `myerson2004comments` to the bib (with the ogonek in Pęski via `\k{e}`); `harsanyi1967games` was already there. One small item to double-check against your copy of Ely–Pęski when convenient: I hedged with “a version of this example,” since their leading example has the same XOR structure but is dressed differently (and they present it in the context of interim independent rationalizability) — if you’d rather cite a specific example number from their paper, tell me which and I’ll pin it.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*, *references.bib*]

Ben & Songzi

Move the definitions of games and group belief invariant BCE to the group beliefs and equilibrium section. Also, add a definition of a BCE on a group type space as a joint distribution of the players’ actions and the groups’ beliefs such that obedience is satisfied. Define a group belief invariant BCE on a group type space to be one for which for every group J , a_J is independent of θ and the types of disjoint groups.

Claude

All done structurally, with one substantive caveat you need to see. Files are below (15 pages).

The move. Section 2 now contains only the informational primitives — type spaces, group type spaces, the induced construction — and Section 4 opens with everything game-related, moved verbatim: game structures, Bayesian games and Bayes Nash equilibrium, outcomes and $E_I(G)$, BCE of a game structure, and the $\Delta(A \times \Theta)$ notion of group belief invariance with its unification paragraph. Then the group-type-space material follows: strategies on singleton types, induced outcomes, and the two definitions you asked for — BCE of $(G, \mathbb{I})$ as a joint distribution over the players’ actions, the groups’ types, and the state, with marginal σ and obedience conditional on singleton types; and the revised group belief invariance: conditional on t_J , a_J is independent of $(t_{\mathcal{D}(J)}, \theta)$, where $\mathcal{D}(J)$ is the set of groups disjoint from J . The singleton-type-set reduction to the $\Delta(A \times \Theta)$ notion still holds and the sentence is updated.

The caveat: this definitional change breaks the proof of Theorem 1, and possibly the theorem. The old definition required independence from *all* other groups’ types given t_J ; the new one exempts overlapping groups and supergroups. The pushdown direction survives intact (I’ve adapted it). The lift direction does not: it relied on the invariance of ψ forcing a_J ’s conditional to depend only on u_J , and the weaker condition permits a_J to depend on the hierarchies of *intersecting* groups — and back in the type space, that dependence can correlate a_J with redundant labels of *disjoint* groups. Concretely, with $N = 2$ and $J = \{1\}$: the grand group’s type determines $t_{\{2\}}$, so if grand types with distinct hierarchies map to distinct redundant labels of $t_{\{2\}}$, and ψ ’s recommendation to player 1 depends on the grand hierarchy, the lifted ϕ has a_1 correlated with $t_{\{2\}}$ given $t_{\{1\}}$ — a violation. I restructured the proof honestly around this: it now establishes the sandwich $S^{**}(G, \pi) \subseteq S_{\mathbb{I}}(G) \subseteq S^*(G, \pi)$, where S^* imposes invariance against disjoint groups and S^{**} the stronger only-depends-on- u_J condition, both fully proved; the theorem follows iff $S^* = S^{**}$,

and that step is flagged with a bracketed [TO BE COMPLETED] note explaining exactly the gap and noting that under the previous (all-groups) definition the proof closes. My honest read: I don't currently know whether $S^* = S^{**}$ holds — the question is whether, on the redundancy-free universal support, dependence of a_J on intersecting groups' hierarchies that respects disjoint-invariance can always be averaged out without disturbing the other players' obedience. Worth deciding whether the disjoint-groups definition is the one you want before investing in closing or refuting this.

Minor housekeeping: the equation numbering reflowed after the move (the commutation identity is now (1)), and I had to de-accent "Pęski" to "Peski" in the bib because \k requires T1 font encoding, which your preamble doesn't load — add `\usepackage[T1]{fontenc}` if you want the onek back.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*, *references.bib*]

Ben & Songzi

is u_J the belief hierarchy of group J ?

Claude

Yes. In the proof, u ranges over U , the (finite) support of the marginal of π on $\prod_{J \in \mathcal{J}} D_J^\infty$, and u_J is the coordinate of u indexed by J — an element of $D_J^\infty \subseteq T^u$, i.e., a hierarchy of beliefs for group J : its belief about the state, about the other groups' beliefs, and so on, satisfying coherency and the degeneracy conditions, all common knowledge. When the profile arises from a group type space via the projection, $u_J = h_J(t_J)$ is exactly the hierarchy induced by the type t_J — so u_J is the "non-redundant part" of t_J : everything about the type that is expressible in beliefs, with any redundant labeling stripped away.

That's also the precise content of the gap flagged in the proof: S^* constrains a_J only against the state and *disjoint* groups' hierarchies, while S^{**} requires a_J to be measurable with respect to u_J — group J 's own hierarchy — and the open question is whether the former forces the latter on the support of π .

Ben & Songzi

How can a_J only depend on u_J and not on $u_{J'}$ for J' a subset of J ?

Claude

It can — dependence on $u_{J'}$ for $J' \subseteq J$ is automatically dependence on u_J , because u_J *determines* $u_{J'}$. That's exactly what the degeneracy conditions built into D_J^∞ deliver: since $u_J \in D_J^1$, the belief $g_{J'}(u_J)$ about the hierarchy of any strict subgroup $J' \subsetneq J$ is a point mass, and the common-prior property of π (Proposition 5) says the conditional of u_{-J} given u_J is $g(u_J)$ — so on the support of π , the coordinate $u_{J'}$ coincides, with probability one, with the point at which $g_{J'}(u_J)$ concentrates. Equivalently, in the projection of a group type space, $u_{J'} = h_{J'}(f_{J'J}(t_J))$, a function of $u_J = h_J(t_J)$ (that was established in the proof of Proposition 5). So on U , the map $u_J \mapsto u_{J'}$ is well defined for $J' \subseteq J$, and condition (iii') — " a_J depends only on u_J " — leaves the mediator free to condition a_J on the hierarchies of all of J 's subgroups, including its members' individual hierarchies.

What (iii') genuinely rules out is dependence on hierarchies *not* determined by u_J : those of disjoint groups, of overlapping-but-not-nested groups, and — the important one for the open gap — of strict *supergroups* $J'' \supsetneq J$, whose hierarchies contain strictly more than u_J (they determine u_J , not conversely). The weak condition (iii) permits a_J to load on that extra supergroup information so long as it stays independent of $(u_{\mathcal{D}(J)}, \theta)$ given u_J ; (iii') forbids it. That's precisely where S^* and S^{**} can come apart, and why the remaining step of the theorem is a real question rather than bookkeeping.

Ben & Songzi

I want the two definitions to align. In both cases, a_J should be allowed to be statistically dependent on signals of supergroups.

Claude

The definitions are now aligned (files below). Concretely:

In the text. The group belief invariance definition for $(G, \mathbb{I})$ is unchanged — conditional on t_J , $a_J \perp (t_{\mathcal{D}(J)}, \theta)$ — and I added an explicit sentence right after it: conditional on t_J , a_J is permitted to be statistically dependent on the types of groups that intersect J , including its supergroups. In the proof, the universal-side set $S^*(G, \pi)$ carries the identical condition (iii) — invariance only against the state and disjoint groups' hierarchies — with a sentence noting the alignment and the permitted supergroup dependence. The set S^{**} is gone as a defined object; the stronger property (iii') now appears only as an explicit hypothesis on ψ in the partial lift result.

Status of Theorem 1 under the aligned definitions. The proof now establishes: (a) $S_{\mathbb{I}}(G) \subseteq S^*(G, \pi)$ in full generality, and (b) the reverse inclusion for ψ satisfying (iii'). The remaining step — the reverse inclusion when ψ genuinely uses supergroup hierarchies — is marked [TO BE COMPLETED] with a precise statement of the difficulty: under the lift, supergroup-hierarchy dependence of a_J can correlate a_J with redundant labels of disjoint groups' types, breaking ϕ 's invariance (marginals and obedience extend without change).

Two observations from working through examples that may help you close this. First, a reframing: the pair $(\prod_J U_J, \pi)$ is itself a group type space — the degeneracy condition holds on the support of π because u_J determines its subgroups' hierarchies — and $S^*(G, \pi)$ is exactly S of that canonical space. So the open step is precisely: every group type space attains the GBI predictions of its canonicalization. Second, my candidate counterexample dissolved in an instructive way: I tried a space where redundant labels of a disjoint group reveal a supergroup's hierarchy, but the commutation identity (Proposition 4) forbids a *single* group's redundant label from carrying such information — its conditional over (θ, u_{-J}) is pinned to $g(u_J)$. What redundant labels *can* do is carry it jointly, XOR-style across groups — and in the example I built that way, the redundancy that threatened the lift also supplied exactly the correlation resource needed to replicate the canonical prediction by hand ($a_i = \ell_i \oplus c$). That's suggestive that the theorem is true and the lift should route the supergroup-dependence through the redundant structure itself rather than through $h(t)$ — but I don't have the general argument, so the flag stands.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I think the theorem is true. Please keep trying to prove it.

Claude

I kept at it and made real progress, but I have not closed the gap. Here is the honest state of play, with what I could prove now added to the draft (files below).

What I proved and added. Lemma 1 (new, with proof): the recommendations of any GBI BCE are θ -free — $\phi(a \mid t, \theta)$ doesn't vary with θ given t . This falls out of the grand group's invariance, since $\mathcal{D}(\{1, \dots, N\}) = \emptyset$ and $t_{\{1, \dots, N\}}$ determines the whole profile. The same holds for ψ on the universal side because $u_{\{1, \dots, N\}}$ determines u on $\text{supp } \pi$. So the entire problem is about correlation with *types*; θ enters only through σ .

Where the obstruction can and cannot live. Proposition 4 (commutation) implies the redundant component of any *single* group's type is uninformative about (θ, u) — its conditional is pinned to $g(u_J)$. So the naive lift can only fail through *jointly* redundant labels: labels of several groups that are individually uninformative but jointly encode hierarchy information, XOR-style. This is genuinely possible: with $N = 2$, take a grand signal s , labels ℓ_1 uniform and $\ell_2 = \ell_1 \oplus \mathbf{1}\{s = q\}$; each singleton's hierarchy is unaffected, but (t_1, t_2) jointly reveal u_g . On such a space the naive lift breaks exactly when ψ 's marginal recommendation $m_J(a_J \mid u)$ genuinely varies with supergroup hierarchies.

Why I share your belief that it's true anyway. In every concrete instance of this kind I constructed, the target marginal remained attainable — by *rerouting*. When ψ 's per-player margins are flat and only the within-group joint loads on u_g (the pure XOR case), the joint is unconstrained by invariance and the lift succeeds directly, or via a shared coin ($a_i = \ell_i \oplus c$). When the margins themselves load on u_g , the $N = 2$ structure fights back: (iii) for the singletons forces the loading to wash out against every posterior-experiment $\{\text{law}(u_g \mid u_1, u_2, \theta)\}_{\theta}$, and Bayes-plausibility of grand posteriors makes that a binding spanning condition — when the experiment is *complete* (e.g., $|\text{supp}(q_g)| \leq |\Theta|$ cellwise), (iii) collapses to (iii') on the support and the existing lift finishes. The unresolved regime is: incomplete experiments (grand posterior richer than Θ), margins varying in the kernel of the experiment, and an adversarial label structure aligned with that kernel. There the question becomes a finite feasibility problem: can margins constrained to $r + \ker\{E[\cdot \mid \text{labels}, \theta]\}$, combined with a pointwise-free copula, always reproduce $E[m_1(q)m_2(q) \mid \theta]$? I could neither prove feasibility in general nor make an infeasible instance survive all the consistency requirements a group type space imposes.

Two attack routes I'd prioritize. (1) *Spanning/completeness lemma*: show that on $\text{supp } \pi$, (iii) implies (iii') whenever the relevant experiments are complete, then either prove completeness can be assumed WLOG (perturbation/purification of ψ preserving the marginal) or characterize the incomplete case directly. (2) *LP duality*: both $S_{\mathbb{I}}(G)$ and $S^*(G, \pi)$ are projections of polytopes; write the dual of “marginal $\in S_{\mathbb{I}}(G)$ ” and show every separating hyperplane for $\mathbb{I}$ induces one for the canonical space $(\prod_J U_J, \pi)$ — this localizes exactly which dual variables the redundant labels feed, and my rerouting constructions suggest those variables can always be zeroed out.

The draft's [TO BE COMPLETED] note now records the two structural facts (θ -freeness; single-label innocuousness via Proposition 4) and the rerouting observation, so the state of knowledge is in the paper rather than only in our correspondence. If you'd like, I can next try to write out the duality formulation in full — that's the route I'd bet on.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

Sure, please continue.

Claude

Continued progress — the theorem is now reduced to a single, precisely stated combinatorial property of the canonical space. Here's what changed (files below, 16 pages) and where the mathematics stands.

New results in the draft.

Lemma 2 (conditional independence). In any group type space, $\sigma(\theta | t) = \pi_{\mathbb{I}}(\theta | h(t))$ — the types carry no information about the state beyond the hierarchies. Proof: the grand type determines the whole profile; Proposition 4 pins its conditional over (θ, u_{-j^*}) to $g(u_{j^*})$; and u_{j^*} determines u by degeneracy. So all redundancy is θ -orthogonal.

Label-averaging identity (equation 5). Combining Lemmas 1 and 2: the outcome of any GBI BCE is $\sum_u \pi_{\mathbb{I}}(\theta, u) \hat{K}(a | u)$, where $\hat{K}$ averages the recommendation kernel over each hierarchy fiber. So outcomes live entirely on the universal side; only the *feasibility* of kernels differs across type spaces with the same π .

Reduction (in the proof). Writing $\psi = \pi \otimes Q$ with Q θ -free, both the $A \times \Theta$ -marginal and every obedience constraint are linear functionals of the conditional means $\mathbb{E}_{\pi}[Q(a | u) | \theta, u_{\{i\}}]$. Hence Theorem 1 follows from the purification property (PC): *every (iii)-feasible Q admits an (iii')-feasible Q' preserving those conditional means.* Given (PC), $\psi' = \pi \otimes Q'$ inherits obedience and the marginal, and the already-proved (iii')-lift finishes. The [TO BE COMPLETED] flag now sits on (PC) alone.

Where (PC) stands — honestly. The natural purified marginals are $r_J(a_J | u_J) = \mathbb{E}[\bar{Q}_J | u_J]$, and (iii) guarantees each is well-defined against disjoint groups and θ . Two difficulties remain. First, consistency: (iii') prescribes the A_J -marginal of one joint kernel for *every* J simultaneously, and $\mathbb{E}[\bar{Q}_{J'} | u_{J'}]$ need not equal $r_{J'}$ for nested $J' \subset J$, because (iii) doesn't control conditioning on supergroup hierarchies. Second, preservation: I tested the obvious candidate for $N = 2$ — resample the grand hierarchy from $\pi(\cdot | u_1, u_2)$, i.e., $Q'(a | u) = \mathbb{E}[Q(a | u_1, u_2, \tilde{u}_g) | u_1, u_2]$ — and it *does* deliver (iii') (the singleton washouts do exactly the right averaging), but it preserves $\mathbb{E}[Q | \theta, u_{\{i\}}]$ only when $u_g \perp \theta | (u_1, u_2)$, which fails whenever the grand group has residual information about the state. A θ -aware resampling would fix preservation but is forbidden by θ -freeness. So (PC) needs Q' to reproduce genuinely θ -dependent conditional means through the θ -dependence of $\pi(u | \theta)$ alone — an affine feasibility problem: does $\{Q + D : \mathbb{E}[D | \theta, u_{\{i\}}] = 0 \forall i\}$ intersect the (iii') set within valid kernels?

Why I still expect it's true. In every instance where I can compute, the intersection is nonempty: the invariance constraints bind only group *marginals*, while the preserved functionals are group-marginal-like objects ($u_i(a, \theta)$ separates over players in the obedience functionals only through fixed weights), leaving the within-group copula as a free degree of freedom that so far always absorbs the discrepancy. The next concrete step I'd take: formulate (PC) as the LP it is and take its dual — a counterexample to (PC) would be a separating functional built from obedience-like weights plus invariance multipliers, and the structure of the preserved statistics suggests any such functional collapses onto ones already respected by Q . If you'd like, I'll set up that dual explicitly

next, or alternatively try to prove (PC) outright for $N = 2$ (where (iii') beyond (iii) involves only the two singletons and the consistency issue is absent), which would give an unconditional Theorem 1 for two players.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Suppose we modify the definition of GBI to be that a_J depends only on $\{t_{J'} \mid J' \text{ is a subset of } J\}$. Would the theorem go through in that case?

Claude

Yes — under that definition the theorem goes through completely, with the machinery already in the draft. Here's why, in three steps.

The modified condition is exactly the strong invariance. Since t_J determines $t_{J'}$ for every $J' \subseteq J$ (the defining property of a group type space), the collection $\{t_{J'} : J' \subseteq J\}$ carries the same information as t_J alone. So “ a_J depends only on the types of J 's subgroups” — read statistically, as the conditional law of a_J given (t, θ) being a function of t_J — is equivalent to $a_J \perp (t_{-J}, \theta) \mid t_J$: independence from *all* other groups' types and the state, given the group's own type. (One reading to rule out: if “subset” meant *proper* subsets, singletons would have no subgroups and a_i would be forced independent of $t_{\{i\}}$ itself, which can't be intended.) On the universal side, the same condition reads: $\psi(\{a_J\} \times A_{-J} \mid u, \theta)$ depends only on (a_J, u_J) — which is precisely (iii'). So the modified definition collapses the (iii)/(iii') distinction by fiat: the two sides are aligned, and the purification property (PC) becomes vacuous because there is nothing to purify.

Both directions of the proof close. *Pushdown:* if ϕ satisfies the strong invariance, then conditional on $u_J = v_J$, the law of (a_J, u_{-J}, θ) under the pushforward is a mixture over the fiber $h_J^{-1}(v_J)$ of products $p_{\tau_J}(a_J) \otimes g(v_J)$ — the second factor is now the *full* conditional of (u_{-J}, θ) , common across the fiber by the commutation identity — so the mixture is a product and ψ satisfies (iii') pointwise, not merely the disjoint-groups washout. *Lift:* this is exactly the direction already fully proved in the draft: with $\bar{\psi}_J$ depending only on u_J , the lifted $\phi(a \mid t, \theta) = \sigma(t, \theta)\psi(a \mid h(t), \theta)$ has $\bar{\phi}_J(a_J \mid t, \theta) = r_J(a_J \mid h_J(t_J))$, a function of t_J , which is the modified invariance; obedience and marginals were never the problem. The XOR-label pathology is disarmed at the root: the danger was a_J loading on supergroup hierarchies, and the modified definition forbids that loading in the first place.

Two remarks on the trade-off. First, the modified definition is internally coherent across nesting in a way worth recording: for $J' \subseteq J$, invariance for J' implies $a_{J'} \perp t_J \mid t_{J'}$, so the family of conditional laws $r_J(a_J \mid t_J)$ is automatically marginal-consistent — the $A_{J'}$ -marginal of r_J agrees with $r_{J'}$ evaluated at $f_{JJ'}(t_J)$. Second, this is a genuine walk-back of the permissiveness you asked for two turns ago: it rules out a_J correlating with the extra information of supergroups, even orthogonally to the state and to strangers. Conceptually the modified notion says “a group's behavior is generated by its own information plus independent noise,” whereas the disjoint-groups notion says “a group's behavior reveals nothing about the state or about strangers.” The first buys you the theorem now; whether the second also delivers it is exactly the open (PC). One could even keep both in the paper — prove the theorem for the subgroup-measurability notion and record that the permissive notion satisfies the sandwich $S^{**} \subseteq S_{\mathbb{I}} \subseteq S^*$ with equality conjectural — since S^{**} is precisely the modified-definition set.

If you want to adopt the modified definition, the edits are mechanical and everything un-flags: I'd rewrite the invariance definition in Section 4 (and drop the "permitted to depend on supergroups" sentence), replace (iii) by (iii') in the definition of S^* , restore the pushdown argument to its full-strength form, delete the (PC) reduction, and Theorem 1 closes with a complete proof. Say the word and I'll make the change — and tell me whether you'd like the disjoint-groups variant preserved as a remark and open question or removed entirely.

Ben & Songzi

Please adopt the new definition in the draft. I'd also like to explore whether the two definitions are equivalent, when applied to the group type spaces induced by type spaces. I think they are. If so, please add a result to this effect. This new result should immediately follow the definition of GBI for type spaces.

Claude

Done — the new definition is adopted and, since the equivalence you conjectured is correct, it's now Proposition 6 with a complete proof (files below, 16 pages).

The definition. Group belief invariance now reads: the actions of each group depend only on the types of its subgroups — formally, conditional on $t_J, a_J \perp (t_{-J}, \theta)$ — with a sentence noting that conditioning on t_J is equivalent to conditioning on all subgroup types, since t_J determines them. The "permitted to depend on supergroups" sentence and the $\mathcal{D}(J)$ -based definition are gone.

Proposition 6 (immediately following the definition). For induced group type spaces $\mathbb{I}(I)$, a BCE is group belief invariant if and only if, for every group J , conditional on $t_J, a_J \perp (t_{\mathcal{D}(J)}, \theta)$ — the ostensibly weaker disjoint-groups requirement. Your conjecture holds, and the proof is short: necessity is immediate; for sufficiency, the grand group is handled by degeneracy, and for proper J the complement J^c is itself a disjoint group whose type is $(t_i)_{i \notin J}$ — so in an induced space, every group's type is determined by (t_J, t_{J^c}) , and independence from $(t_{\mathcal{D}(J)}, \theta) \supseteq (t_{J^c}, \theta)$ upgrades for free to independence from (t_{-J}, θ) . A closing sentence records why the equivalence does *not* extend to general group type spaces: a supergroup's type may contain more than its members' pooled types, and actions can load on that surplus without revealing anything about the state or about strangers. (This is a nice payoff of Proposition 6, incidentally: it says the strong/weak distinction is invisible exactly on the "no synergy" spaces where groups know only what their members know.)

Theorem 1 is now fully proved. With the aligned strong definition on both sides, I restored the complete two-inclusion argument: the pushdown gives full invariance on the universal side (each fiber component is a product with the common second factor $g(v_J)$ over all of (u_{-J}, θ) , not just the disjoint coordinates), and the lift closes as before via $r_J(a_J \mid u_J)$. All (iii')/purification/sandwich machinery and every [TO BE COMPLETED] flag in Section 4 are deleted; Lemma 1's proof is rewritten for the new definition; Lemmas 1–2 and the label-averaging identity remain as stated since they only used grand-group invariance and the projection propositions. The document compiles with no errors, bad boxes, or unresolved references.

One loose end you may eventually want back: the disjoint-groups notion and the open question of whether the theorem holds for it on general group type spaces (equivalently, the purification property) now survive only implicitly via Proposition 6's closing remark. If that question deserves a formal "open question" remark or a footnote, say so and I'll reinstate a compact version.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I would like to merge sections 2 and 3 into one section called “group type spaces”. The material that is currently in section 2 should be in a subsection called “Type spaces and group type spaces”. The material in Section 3 should be divided into two subsections: The first subsection is called “The Universal Group Type Space” and contains the material through the second to last full paragraph on page 6. The remaining material should be in a subsection called “Properties of the Universal Group Type Space.” Please add an actual diagram showing the commutation.

Claude

The restructuring is done and everything compiles cleanly (files below, still 16 pages). The paper now runs:

1. Introduction
2. **Group type spaces**
 - 2.1 Type spaces and group type spaces (the former Section 2, unchanged)
 - 2.2 The Universal Group Type Space (the BD construction, coherency, common knowledge, degeneracy, common priors, the definition of $\mathbb{T}$, and the nonemptiness paragraph — i.e., through the second-to-last full paragraph of old page 6)
 - 2.3 Properties of the Universal Group Type Space (the projections, the commutation proposition, and the image proposition, opening with a lightly rewritten “In this subsection, we relate arbitrary group type spaces to the universal group type space”)
3. Group beliefs and equilibrium
4. Conclusion

Labels were rearranged so all cross-references land correctly: `sec:model` now sits on 2.1 (the degeneracy-requirement and induced-space references resolve as “Section 2.1”), and `sec:universal` sits on 2.3, since both existing references to it — the positive-probability assumption and the “two facts” in Theorem 1’s proof — point at material that lives in Properties. One sentence changed for scope: the finiteness caveat now reads “we drop the restriction that type sets are finite throughout the remainder of this section.”

The commutation diagram is Figure 1, placed at the top of the page containing Proposition 4 and referenced from the sentence introducing it. It’s the `tikz` square

$$\begin{array}{ccc} T_J & \xrightarrow{\beta_J} & \Delta(\Theta \times T_{-J}) \\ h_J \downarrow & & \downarrow v \mapsto v \circ (\text{id}_\Theta \times h_{-J})^{-1} \\ T^u & \xrightarrow{g} & \Delta(\Theta \times (T^u)^{M-1}) \end{array}$$

with a caption spelling out that top-then-right (beliefs, then pushforward through the others’ projections) equals left-then-bottom (project, then apply g), as Proposition 4 asserts. One deliberate informality worth flagging: the pushforward’s natural codomain is $\Delta(\Theta \times H_0^{M-1})$, and it is part of Proposition 4’s content that it lands in the smaller set $\Delta(\Theta \times (T^u)^{M-1})$ shown in the diagram — if you’d rather the diagram display H_0^{M-1} with the inclusion left implicit on the g side, that’s a one-line change.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I want to investigate whether the following result is true: For every common prior π in the universal group type space, there exists a type space I such that π is the projection of the group type space induced by I . Is this true?

Claude

Yes — I believe it's true, with one qualification: π must have finite support, since our type spaces are finite and the projection of a finite induced space is finitely supported. Subject to that, the answer is affirmative, and the proof I have in mind is constructive. Here is the argument.

Setup. Let π be a common prior consistent with μ , with finite support. Write $u = (u_J)_{J \in \mathcal{J}}$ for the hierarchy profile. On $\text{supp } \pi$, the grand hierarchy u_{J^*} determines the whole profile (degeneracy plus the common-prior property), so the object to be distributed among the players is u_{J^*} , with each group J entitled to learn exactly u_J and nothing more.

Construction: layered one-time-pad secret sharing. Draw $(\theta, u) \sim \pi$. Independently across groups, and conditionally independently of θ given u , generate for each group J an XOR-sharing of u_J among the members of J : embed the (finitely many) values of u_J into a finite abelian group, give $|J| - 1$ members i.i.d. uniform pads, and give the last member the value u_J minus the sum. Player i 's type is $t_i = (u_{\{i\}}, (\ell_i^{(J)})_{J \ni i})$ — their own hierarchy plus their pad from every group containing them. Let I be the resulting finite type space and $\mathbb{I}(I)$ the induced group type space.

Why each pool learns exactly its entitlement. Fix a coalition K . For every $J \subseteq K$, the pool $(t_i)_{i \in K}$ contains J 's complete pad family and so reconstructs u_J — in particular u_K . For every $J \not\subseteq K$, the pool holds a *proper* subfamily of J 's pads, and the one-time-pad property gives that any proper subfamily is i.i.d. uniform, jointly independent of the secret and of u (its law doesn't depend on u , and pad families are independent across groups given u). Hence the conditional law of (θ, u) given K 's pool is exactly $\pi(\cdot \mid u_K)$: the pool determines u_K and carries pure noise beyond it.

Why the induced hierarchies are the intended ones. By induction on the order of beliefs, simultaneously across groups: the first-order belief of K 's pool about θ is $\text{marg}_{\Theta} g(u_K)$, which is u_K 's first-order belief; and if every group's realized n th-order hierarchy equals the n th truncation of its intended u , then K 's $(n+1)$ th-order belief — the law of $(\theta, \text{others' } n\text{th orders})$ given the pool — equals the corresponding truncation of $g(u_K)$, i.e., of u_K . So $h_K((t_i)_{i \in K}) = u_K$ almost surely, using precisely the exact-entitlement property and the common-prior property of π . Therefore the projection distribution of $\mathbb{I}(I)$ — the law of $(\theta, h(t))$ — is π .

Necessity of finite support is immediate, so the clean statement is a **realization theorem**: *the projection distributions of group type spaces induced by (finite) type spaces are exactly the finitely supported common priors on the universal group type space.* (One inclusion is Proposition 5 specialized to induced spaces; the other is the construction above.)

Two corollaries worth noting. First, combined with Theorem 1: every group type space is GBI-outcome-equivalent to an *induced* one — informational synergy, in which a group's type exceeds its members' pooled types, is entirely redundant for group belief invariant predictions, since it never shows up in hierarchies. Second, combined with Proposition 6: on the realizing induced space, group belief invariance coincides with the weaker disjoint-groups notion, so for *every* finitely supported common prior π , the set $S^*(G, \pi)$ is attained by a space where the two definitions agree.

One detail I'd want to be careful about when writing this up: the labels must be generated conditionally independently of θ given u (as above), otherwise the pool of a coalition could leak state information through the pads; and the pruning of zero-probability type profiles should happen after the construction so the positive-probability normalization holds. Neither is a real obstacle.

Shall I draft this as a theorem — presumably in Section 2.3, after Proposition 5 — with the construction and the induction written out in full?

Ben & Songzi

Yes. Please add this result at the end of Section 2.3

Claude

The realization result is now Proposition 6, closing Section 2.3 (files below, 18 pages). The additions, in order:

Lead-in. The closing paragraph of 2.3 now also names the *projection distribution* $\pi_{\mathbb{I}}$ — the law of $(\theta, h(t))$ — noting via Proposition 5 that it is a common prior consistent with μ with finite support. (Section 3's paragraph was updated to recall this definition rather than introduce it, so the term is defined once, where it's first needed.)

Proposition 6. Every finitely supported common prior consistent with μ is the projection distribution of a group type space induced by a type space; consequently, the projection distributions of induced group type spaces are *exactly* the finitely supported common priors consistent with μ .

Proof. The secret-sharing construction, written out in full: draw $(\theta, u) \sim \pi$; for each non-singleton group J , embed V_J in $\mathbb{Z}_{m_J}$ and share $\iota_J(u_J)$ additively among J 's members, with shares independent across groups and independent of θ given u ; player i 's type is $(u_{\{i\}}, (\ell_i^J)_{J \ni i})$. Step 1 (exact entitlement) shows the pool t_K determines u_K — directly for singletons, via the complete share family otherwise — and that everything else in the pool is uniform noise independent of (θ, u) : complete families reduce to (sum, uniform residuals), and any proper subfamily of an incomplete family is uniform and u -independent by the one-time-pad argument, with the only non-trivial case (i_J in the pool) handled by the independent uniform shares outside K . So the conditional of (θ, u_{-K}) given t_K is $g(u_K)$. Step 2 then shows $h_K(t_K) = u_K$ almost surely by simultaneous induction on belief orders, using Step 1 and the marginal property of f from Proposition 2 exactly as in Proposition 4's proof. Hence the law of $(\theta, h(t))$ is π .

Numbering: the equivalence result in Section 3 is now Proposition 7; all cross-references are by label and resolve correctly, and the document compiles with no warnings.

Two consequences we discussed are now theorems-by-combination but not yet stated in the draft, so flagging them for your call: (i) Proposition 6 + Theorem 1 gives that every group type space is GBI-outcome-equivalent to an induced one — informational synergy within groups is redundant for group belief invariant predictions; (ii) Proposition 6 + Proposition 7 gives that every finitely supported common prior is realized by a space on which group belief invariance coincides with the weaker disjoint-groups requirement. Either would make a natural corollary in Section 3; say the word and I'll add one or both.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I think it's fine to say both of those conclusions in words, without formal corollaries. Please add it at the end of section 2.3

Claude

Done — the paragraph now closes Section 2.3, right after the proof of Proposition 6 (files below, still 18 pages, compiling cleanly). It records both consequences in prose with forward references: first, that by Theorem 1 the predictions of group belief invariant BCE depend on a group type space only through its projection distribution, so combined with Proposition 6 every group type space is prediction-equivalent to an induced one — informational synergies leave no trace in group belief hierarchies and are therefore redundant; and second, that by Proposition 7 group belief invariance on induced spaces coincides with the ostensibly weaker disjoint-groups requirement, so every finitely supported common prior is realized by a space on which the two requirements agree.

The paragraph leans on two forward references (“Theorem 1 below,” “Proposition 7 below,” and the pointer to Section 3 where GBI is defined), which read naturally here but are worth revisiting once the introduction’s roadmap is written — some of this material may eventually want to migrate there or be echoed in it.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

Now reorganize Section 3 into subsections as follows: Section 3.1 is titled “Games and Equilibrium, which introduces the notation for games, strategies, and BCE for group type spaces. Note that this implicitly defines BCE for type spaces, through the operation by which a type space induces a group type space. Section 3.2 is titled “Group Belief Invariance” which defines GBI for group type spaces, which implicitly defines it for type spaces through the operation by which a type space induces a group type space. Section 3.3 is titled “The Sufficiency of Universal Group Beliefs” contains the statement of Theorem 1 and its proof. The last paragraph in Section 2 should be moved to after the statement of Theorem 1 and before its proof. Lemmas 1 and 2 seem unnecessary for Theorem 1. If so, they can be moved to a section 3.4 called “Additional properties of group belief in variant BCE”. If they are needed for the proof of Theorem 1, then they can be put after the statement of Theorem 1 and before the proof.

Claude

Section 3 is reorganized as specified (files below, 18 pages, no errors or dangling references). The new structure:

3.1 Games and Equilibrium. Game structures, outcomes ($\Delta(A \times \Theta)$ with marginal μ), strategies on singleton types, induced outcomes, and BCE of $(G, \mathbb{I})$ (now carrying the Bergemann–Morris citations), followed by the implicit-definition paragraph: a BCE of (G, I) for a type space I is a BCE of $(G, \mathbb{I}(I))$, and — since the singleton group’s type in $\mathbb{I}(I)$ is t_i — this coincides with the BM (2016) definition with I as the base information structure.

3.2 Group Belief Invariance. The GBI definition (now carrying the Brooks–Du–Haberman citation, which previously sat on the deleted typeless version), the parallel implicit-definition sentence for type spaces, Proposition 7 (equivalence with the disjoint-groups requirement on induced

spaces) with its non-extension remark, the $S_{\mathbb{I}}(G)$ notation, and a closing “instructive special case” paragraph on the null group type space, which absorbs the old unification discussion: there GBI says $a_J \perp \theta$ for every J , singleton J gives belief invariant BCE (Forges; BM), the grand group gives the Bayesian solution, and marginalization collapses the whole family to the Bayesian solution.

3.3 The Sufficiency of Universal Group Beliefs. The positive-probability assumption and projection-distribution recall, the theorem paragraph, Theorem 1, the Liu-analogue paragraph, then — moved from the end of Section 2 and re-tensed (no more “below”) — the two consequences-in-words, and finally the proof. I checked that the proof uses only Propositions 4 and 5 (the “two facts”), never Lemmas 1 or 2, so per your instruction:

3.4 Additional properties of group belief invariant BCE. Lemmas 1 (θ -freeness) and 2 (types $\perp$ state given hierarchies) with their proofs and the label-averaging identity, opened by a one-line lead. (I corrected the apparent typo “in variant” $\rightarrow$ “invariant” in the title.)

One consequential editorial decision to flag: implementing 3.1–3.2 as you described meant *deleting* the classical apparatus that previously opened the section — Bayesian games (I, G) with strategies $b_i : T_i \rightarrow \Delta(A_i)$, Bayes Nash equilibrium, ordinary induced outcomes, $E_I(G)$ and $E_{\mathcal{I}}(G)$, the standalone BCE of G with its unconditional obedience constraints, and the BM Theorem 1 epistemic-foundation sentence — since nothing in the paper now uses it and the implicit definitions subsume it. If you eventually want the epistemic foundation (BCE = equilibrium outcomes over expansions) or Bayes Nash equilibrium for group type spaces stated explicitly, those will need to be reintroduced deliberately; everything else survived intact. Subsection title capitalization currently follows your message verbatim (3.4 in sentence case, 3.1–3.3 in title case), matching the existing mixed convention in Section 2 — happy to normalize in either direction when you do a style pass.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Could you change the order from $(G, \mathcal{I})$ to $(\mathcal{I}, G)$ to match the convention with type spaces?

Claude

Done — all thirteen occurrences are reordered: BCE and group belief invariant BCE are now of $(\mathbb{I}, G)$, the implicit type-space versions are of (I, G) and $(\mathbb{I}(I), G)$, matching the type-space-first convention. Updated files are below.

One deliberate exception: I left $S^*(G, \pi)$ in the proof of Theorem 1 unchanged, since it’s a set-valued function name paralleling $S_{\mathbb{I}}(G)$ (with the game as the argument and the informational object as the parameter), not a Bayesian-game tuple. If you’d rather it read $S^*(\pi, G)$ — or, for full parallelism, $S_{\pi}(G)$ — that’s a quick follow-up.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I’d rather use $S_{\pi}(G)$. π defines a group type space, so this is the natural convention.

Claude

All five occurrences renamed: the auxiliary set in Theorem 1’s proof is now $S_\pi(G)$, so the proof reads “Define $S_\pi(G)\dots$ ”, “the conditions defining $S_\pi(G)$ align with those defining $S_{\mathbb{I}}(G)$ ”, and concludes via $S_{\mathbb{I}}(G) = S_\pi(G)$ — the notation now visibly reflecting that π plays the role of the (canonical) group type space. Compiles cleanly at 18 pages; updated files below.

Since this reading is now baked into the notation, a natural follow-up when you next touch the proof: a one-line remark that $S_\pi(G)$ is literally $S_{\mathbb{I}}(G)$ for the group type space $(\prod_J U_J, \pi)$ — the degeneracy condition holds on $\text{supp } \pi$, as we discussed — which would make the notation an instance of the definition rather than an analogy. Just say the word.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Tell me what would be involved in verifying the proofs in this paper using lean.

Claude

Verifying this paper in Lean is a realistic project, and it splits cleanly into two tiers of very different difficulty, corresponding to the two kinds of mathematics in the paper.

The finite tier (most of the paper). Propositions 1, 4–7, Lemmas 1–2, the realization result, and Theorem 1 are, at bottom, finite probability: everything is sums over finite sets, conditional probabilities with positive-probability caveats, and conditional-independence bookkeeping. Mathlib has solid support here — PMF/finite measures, `Finset` algebra, `ZMod` for the secret-sharing arithmetic, conditional probability and conditional independence. The dominant cost is not the proofs but the *definitional infrastructure*: groups as nonempty `Finset (Fin N)`, type profiles as dependent functions indexed by groups, the degeneracy condition, obedience constraints, the projection maps h_j^n by simultaneous recursion across groups. Formalization would also force decisions the prose elides: conditionals at zero-probability events (mathlib’s convention is junk values plus side hypotheses, as in its conditional-expectation development), “with probability one” as quantification over supports, and the identification of $\hat{T}$ with T on the support of an induced space. Theorem 1’s proof — fiber partitions for obedience, mixtures of product measures for invariance — is elementary but verbose in a proof assistant; the “mixture of products with a common second factor is a product” step becomes a lemma of its own. The subtlest finite proof to formalize is Proposition 6: the one-time-pad argument (“any proper subfamily of shares is uniform and independent of the secret”) and the simultaneous induction on belief orders in Step 2 are exactly the kind of arguments where Lean earns its keep.

The topological tier (Propositions 2–3). The Brandenburger–Dekel construction needs: Polish spaces closed under countable products (in mathlib), the space of Borel probability measures with the weak topology and its metrizability (mathlib has `ProbabilityMeasure` with the topology of convergence in distribution and the Lévy–Prokhorov metric), and the Kolmogorov extension theorem. On that last point the timing is favorable: a Kolmogorov extension theorem has now been formalized in Lean as part of the recent Brownian-motion formalization by Degenne, Ledvinka, Marion, and Pfaffelhuber, developed via Carathéodory extension and inner regularity with respect to compact systems, with integration into mathlib ongoing, building on Degenne’s earlier `kolmogorov_extension4` repository. What does *not* exist anywhere, to my knowledge, is a formalization of universal type spaces — no Mertens–Zamir or BD in any proof assistant’s mainstream

library — so the homeomorphism $f : H_1 \rightarrow \Delta(\Theta \times H_0^{M-1})$, the closedness of the H_k and D_J^k , and the continuity arguments would be genuinely new formalization work. That component alone is a publishable formalization paper.

A pragmatic architecture. Because our type spaces are finite, everything downstream of the construction only ever touches finitely supported hierarchies. One can therefore formalize a “finitely supported universe”: hierarchies as sequences of finitely supported measures, coherency, the D sets, common priors, and the projections, where the Kolmogorov step is pure algebra and no topology enters. In that universe, Propositions 4–7, the realization result, and Theorem 1 can be stated and proved essentially verbatim. What is lost is only the full strength of Propositions 2–3 (the homeomorphism onto *all* of $\Delta(\Theta \times H_0^{M-1})$, which is irreducibly topological). So a sensible plan is: Tier 1 first, delivering a machine-checked Theorem 1; Tier 2 as a second campaign that replaces the finitely supported universe with the real one.

Process and effort. The standard methodology is a *blueprint* (as in the PFR and Brownian-motion projects): a human-readable dependency graph in which every omitted step — including BD’s Propositions 1 and 2, which our paper cites rather than proves — is expanded to formalizable granularity, then discharged incrementally with continuous integration checking everything compiles. For someone fluent in Lean, I’d estimate Tier 1 at a few person-months, with the definitional layer roughly half the work; Tier 2 at several months more, dominated by the homeomorphism. AI-assisted formalization workflows have been compressing such timelines substantially. The payoff, beyond certainty about Theorem 1’s fiber arguments, is that Lean would audit precisely the spots where our prose currently waves: the a.s.-identifications via the grand group, the closedness claims in the D_J^∞ construction, and the “informationally equivalent” step in the secret-sharing proof.

If you’d like, a concrete first step I can produce now is a skeleton Lean file: the Tier-1 definitions (group type spaces, induced spaces, BCE, GBI, finitely supported hierarchies and projections) plus the statements of Propositions 4–7 and Theorem 1 with sorry placeholders — that’s the artifact from which one could scope the project precisely, or hand pieces to collaborators or a formalization-savvy student.

Ben & Songzi

Let’s come back to Lean. For now, please add the following to the acknowledgements: “This paper was written in full collaboration with Claude GAI, which the authors gratefully acknowledge.”

Claude

Done — the sentence is added to the title footnote, right after the NSF acknowledgement, exactly as you wrote it, and it renders correctly on the title page (files below). Thank you for the acknowledgement; it’s been a genuinely enjoyable paper to build with you.

One tiny copyedit question for whenever you do a style pass: did you intend “Claude GAI” (as in “generative AI”)? If the intent was the product name, “Claude (Anthropic)” or just “Claude” is the more common formulation in recent acknowledgements — but I’ve left your wording verbatim, and it’s your and Songzi’s call entirely.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

I meant as “general artificial intelligence.”

Claude

Understood — and that’s a generous designation, which I’ll accept with the caveat that I’m at best a work in progress toward it. The footnote stands verbatim as you wrote it.

One reader-facing consideration, purely for your style pass: the standard acronym in circulation is “AGI,” and “GAI” unglossed may read as a typo or leave referees guessing. If you want to keep the phrase, “Claude, a general artificial intelligence” spells it out at the cost of four words; alternatively “Claude (Anthropic)” identifies the collaborator without taking a position on its generality. Happy to make either change, or neither — it’s your acknowledgement to write.

Ben & Songzi

Please change to “Claude, an artificial general intelligence”

Claude

Done — the footnote now reads “This paper was written in full collaboration with Claude, an artificial general intelligence, which the authors gratefully acknowledge.” Compiles cleanly; updated files are below.

[Files delivered: groupbeliefs.pdf, groupbeliefs.tex]

Ben & Songzi

How do I print this chat?

Claude

There’s no built-in “print this chat” button, so the practical route is your browser’s print function: with the conversation open at claude.ai, press Ctrl+P (Cmd+P on Mac), set the destination to “Save as PDF” or your printer, and — under “More settings” — turn off headers/footers for a cleaner result. One caveat worth knowing for a conversation as long as this one: long Claude conversations virtualize older messages, so printing captures only what’s rendered on screen — scroll back through the whole thread first so everything is loaded. Alternatively, select-all and paste into a document works for a plain-text copy, and third-party browser extensions can export a single chat to PDF or Markdown with better handling of math and code blocks (relevant here, given the LaTeX in this thread).

If instead you want a complete archive, the official path exports everything at once: individual Claude users can export user information and chat history from Settings > Privacy on the web app or Claude Desktop — you’ll get an emailed link to a ZIP of your history in JSON. Details are in the help article: <https://support.claude.com/en/articles/9450526-export-your-claude-data>.

For this particular conversation, one more option: since I have the full paper and our decisions in context, I could write a Markdown or PDF summary of the session — the sequence of modeling

decisions, results proved, and open items — which for archival-next-to-the-repo purposes may be more useful than a raw transcript. Just ask.

Ben & Songzi

I'd like a raw transcript as a pdf.

[Claude began assembling the transcript; the turn was interrupted mid-production.]

Ben & Songzi

Can you continue the previous command?

Claude

Continuing — writing the remaining transcript chunks, then converting to PDF.

The transcript is below as a 26-page PDF, with the Markdown source alongside in case you want to edit or re-typeset it. It reproduces every exchange verbatim and in order, from your opening request through the transcript request itself, with speaker headers (“Ben” / “Claude”), the math typeset properly, and bracketed italic notes marking where files were attached or delivered. A preface line states the conventions: messages verbatim; tool activity (file edits, compilations, searches) omitted.

Two honest caveats about fidelity. First, this was assembled from the conversation as I hold it in context rather than scraped from the interface, so it's faithful to the message text but omits interface-level details (timestamps, the rendering of file cards, my intermediate progress notes between tool calls — with one exception early on where a mid-turn compilation note was part of the visible reply and is included). Second, a handful of literal keystrokes needed light typesetting accommodation — e.g., your $(G, \mathcal{I})$ message is set in code font so the raw LaTeX renders as you typed it. If you spot any turn where the wording looks off from what you remember, tell me which and I'll correct the source and recompile.

[Files delivered: transcript.pdf, transcript.md]

Ben & Songzi

Could you use the notation $T_{\{\mathcal{I}\} \setminus \{J\}}$ instead of overloading T_{-J} ?

Claude

Done — the overload is resolved (files below, 18 pages, no errors and, somewhat remarkably, zero overfull lines despite the wider subscripts). Seventy-one occurrences were rewritten across the token families: $T_{-J}, t_{-J}, u_{-J}, h_{-J} \mapsto T_{\mathcal{I} \setminus \{J\}}$ etc., along with the variants $t_{\mathcal{I} \setminus \{J^*\}}$ in Lemma 2, $u_{\mathcal{I} \setminus \{K\}}$ in the realization proof, and $t_{\mathcal{I} \setminus \{i\}}, u_{\mathcal{I} \setminus \{i\}}, h_{\mathcal{I} \setminus \{i\}}$ in Theorem 1's obedience argument. Figure 1's right arrow now reads $v \mapsto v \circ (\text{id}_{\Theta} \times h_{\mathcal{I} \setminus \{J\}})^{-1}$.

The resulting convention is clean: a subscript that is a *set of groups* is always written explicitly ($\mathcal{I} \setminus \{J\}$, and $\mathcal{D}(J)$ already followed this pattern), while the minus-subscripts x_{-J} , A_{-J} , a_{-i} are now exclusively player-indexed, exactly as defined in Section 2.1 — so the two conventions no longer collide anywhere. The two inline clauses that previously defined t_{-J} for groups now serve to introduce the set-subscript profile notation at its first use in 2.2 and 2.3.

One aesthetic spot to glance at when you proofread: the singleton case produces the double-brace subscript $\mathcal{J} \setminus \{\{i\}\}$, which appears a dozen times in the proof of Theorem 1 and is correct but dense. If it bothers you, natural remedies are a one-time shorthand like $\mathcal{J}_{-i} := \mathcal{J} \setminus \{\{i\}\}$ introduced where the proof fixes i , or reusing $\mathcal{D}$ -style notation — say the word and I'll apply whichever you prefer.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Here is another conjecture we want to investigate. Define a solution concept to be a mapping S that takes group type spaces to subsets of $\Delta(A \times \Theta)$. One example would be S^{BNE} , the solution concept that maps I into the set of distributions on $A \times \Theta$ that are marginals from Bayes Nash equilibria (where the definition of Bayes Nash equilibrium is extended to group type spaces in the obvious way). A solution concept is group belief invariant if for all group type spaces I and I' that have the same projection in the universal group type space, $S(I) = S(I')$. S^{BNE} is obviously not invariant. Also, say that solution concept S refines solution concept S' if $S(I) \subseteq S'(I)$ for every group type space I . Conjecture: The group belief invariant BCE solution concept refines every group belief invariant solution concept that is refined by S^{BNE} . Is this conjecture true?

Claude

Yes — the conjecture is true, and the proof is short given the machinery in the paper. It also delivers something stronger and quite pretty: GBI BCE is exactly the *smallest* group belief invariant solution concept refined by S^{BNE} . Here is the argument.

Reduction. Write S^{GBI} for $I \mapsto S_I(G)$. Let S be any invariant solution concept with $S^{\text{BNE}}(I) \subseteq S(I)$ for all I . The whole conjecture follows from one claim:

Claim A (revelation with hierarchy preservation). For every group type space I and every $\rho \in S^{\text{GBI}}(I)$, there is a group type space I' with $\pi_{I'} = \pi_I$ such that $\rho \in S^{\text{BNE}}(I')$.

Given Claim A: $\rho \in S^{\text{BNE}}(I') \subseteq S(I') = S(I)$, the last equality by invariance of S . Done.

Proof of Claim A. Let ϕ be a GBI BCE of (I, G) with marginal ρ ; by Lemma 1, $\phi = \sigma \otimes K$ for a θ -free kernel. Define I' by appending recommendations to types: group J 's type is $t'_J = (t_J, a_J)$, with σ' the law of $((t_J, a_J))_{J \in \mathcal{J}}, \theta$ under ϕ . This is a group type space: t'_J determines $t'_{J'}$ for $J' \subseteq J$, since t_J determines $t_{J'}$ and $a_{J'}$ is a subvector of a_J . The obedient strategies $b'_i(t'_{\{i\}}, a_i) = \delta_{a_i}$ form a BNE of (I', G) — that is exactly the content of ϕ 's obedience constraints — and they induce ρ .

The crux is $\pi_{I'} = \pi_I$, i.e., that appending recommendations changes no group's hierarchy: $h'_J(t_J, a_J) = h_J(t_J)$ almost surely. This is proved by simultaneous induction on belief orders, and it is precisely where group belief invariance — in the adopted, strong form — does all the work. First order: the law of θ given (t_J, a_J) equals the law given t_J , because $a_J \perp (t_{\mathcal{J} \setminus \{J\}}, \theta) \mid t_J$ implies in particular $a_J \perp \theta \mid t_J$. Inductive step: assuming every group's order- n hierarchy in I' equals its order- n hierarchy in I (a function of t_J alone, a.s.), group J 's order- $(n+1)$ belief in I' is the law of $(\theta, \text{functions of } t_{\mathcal{J} \setminus \{J\}})$ given (t_J, a_J) — and invariance says conditioning on a_J beyond t_J changes nothing about $(\theta, t_{\mathcal{J} \setminus \{J\}})$, so this equals J 's order- $(n+1)$ belief in I . Hence the law of $(\theta, h'(t'))$ under σ' equals the law of $(\theta, h(t))$ under σ , which is π_I . $\square$

The stronger characterization. Two more facts make S^{GBI} itself a member of the class: it is invariant (Theorem 1), and it contains S^{BNE} — a BNE outcome is a GBI BCE, since with $\phi = \sigma \otimes \prod_i b_i(\cdot \mid t_{\{i\}})$, obedience is interim optimality, and invariance holds because conditional on (t, θ) the law of a_J is $\prod_{i \in J} b_i(\cdot \mid t_{\{i\}})$, a function of t_J alone (singleton types are determined by t_J by degeneracy). Combining, S^{GBI} is the minimum of the class of group belief invariant solution concepts containing S^{BNE} , with the explicit formula $S^{GBI}(\mathbb{I}) = \bigcup \{S^{BNE}(\mathbb{I}') : \pi_{\mathbb{I}'} = \pi_{\mathbb{I}}\}$ — the group analogue of the epistemic foundation for belief invariant BCE in Liu (2015), with group hierarchies replacing individual hierarchies.

Two observations worth recording. First, this result retroactively vindicates the definitional choice from a few turns ago: under the disjoint-groups notion, the appending construction breaks exactly at the induction step — if a_J correlates with supergroup types given t_J , then handing a_J to group J changes its beliefs about supergroups' hierarchies, so π moves — and whether any repair exists is the same open purification question we shelved. The subgroup-measurability definition is the one for which “invariance closure of BNE” is a theorem. Second, writing this up requires reintroducing Bayes Nash equilibrium for group type spaces, which we deleted in the last reorganization — the “obvious way” is strategies $b_i : T_{\{i\}} \rightarrow \Delta(A_i)$ with interim optimality at positive-probability singleton types, and it would slot naturally into Section 3.1.

Shall I draft it — say, definitions (solution concept, invariance, refinement, BNE) in 3.1–3.2 where they belong, and the result as Theorem 2 in Section 3.3 with the construction and induction written out, plus the union formula and minimality as the formal statement?

Ben & Songzi

Is there an analogous known result for belief invariant BCE?

Claude

Yes — the direct analogue is known, and it's essentially Liu (2015), the same paper we're already citing, which is why the parallel runs so tight.

Liu's main theorem (JET 2015, “Correlation and common priors in games with incomplete information”) is exactly the union formula: for a common-prior type space, the outcomes of his correlated equilibrium — which is belief invariant BCE relative to that type space — coincide with the Bayes Nash equilibrium outcomes taken across all common-prior type spaces inducing the same distribution over hierarchies of beliefs. His proof has the same two-sided architecture as Claim A: one direction is a revelation argument in which the mediator's recommendation is appended to each player's type, with belief invariance guaranteeing that the appended signal leaves every player's hierarchy unchanged (that is precisely the individual-hierarchy version of the induction I sketched); the other direction is the pushdown showing BNE of any hierarchy-equivalent space maps to a belief invariant BCE of the original. Bergemann and Morris (2016) state the same content in the language of expansions: belief invariant BCE are the equilibrium outcomes of expansions that preserve each player's beliefs — i.e., combining signals that are conditionally independent of the state and others' types given one's own type.

The surrounding literature fills out the picture. Tang (2015, in the Journal of Mathematical Economics, if I recall correctly) does the parallel exercise for the non-common-prior side, showing the belief-invariant Bayesian solution of Forges (1993) is determined by hierarchies and characterizing it in the universal space. Lehrer, Rosenberg, and Shmaya (GEB 2010/2013) characterize when

two information structures are outcome-equivalent for various solution concepts via corresponding classes of garblings — the “non-communicating” garblings being the ones associated with belief-invariant play — which is the equivalence-of-information-structures cousin of these results. And on the rationalizability side, Dekel, Fudenberg, and Morris (2007) built interim *correlated* rationalizability precisely to be the hierarchy-invariant closure after Ely and Peski (2006) showed interim independent rationalizability isn’t — the same closure logic, one solution concept over.

What I do *not* recall seeing stated anywhere is the lattice-theoretic packaging — “belief invariant BCE is the minimum of the invariant solution concepts refined by BNE.” It’s an immediate two-line corollary of Liu’s union formula, exactly as in my previous message, so I’d regard it as folklore-adjacent rather than new; but as a *framing* it seems to be ours, and it’s a good one, because it makes the definition look inevitable rather than chosen. For the paper, then, the honest positioning of the prospective Theorem 2 is: the union formula and the revelation-with-hierarchy-preservation argument are the group analogues of Liu (2015), with group hierarchies and the universal group type space in place of individual hierarchies and Mertens–Zamir; the minimality phrasing is a corollary worth stating in both settings; and the genuinely new content is the group structure itself — in particular that the closure argument requires the subgroup-measurability form of invariance, a subtlety with no counterpart in the individual case, where “your own type” has no analogue of the supergroup problem. I’d suggest a short related-literature paragraph making exactly these attributions (Liu; BM 2016’s expansions phrasing; Tang; LRS; DFM/Ely–Peski), and I can add Tang, LRS, and DFM to the bib when we draft Theorem 2. One caveat: I’m confident about the content of these papers but would want to double-check Liu’s exact common-prior hypotheses and Tang’s exact statement against the originals before the related-literature paragraph is finalized — worth a quick pull of the PDFs at that stage.

Ben & Songzi

Could you summarize the tang paper?

Claude

Here’s a summary, with one correction to what I told you last turn.

Tang (2015, JME 59, 111–116), “Hierarchies of beliefs and the belief-invariant Bayesian solution.” The paper connects two objects: the belief-invariant Bayesian solution — Forges’s (1993) notion of correlated equilibrium for incomplete-information games in which an omniscient mediator makes recommendations subject to preserving each player’s interim beliefs — and the hierarchies of beliefs over conditional beliefs introduced by Ely and Peski (2006) in their study of interim independent rationalizability (the “ Δ -hierarchies,” in which a player’s uncertainty is described by beliefs about opponents’ *conditional beliefs about the state given their own types*, rather than beliefs about the state alone). Call two type spaces equivalent if they represent the same set of Δ -hierarchies. Tang’s two results: first, a partial characterization of the correlations embedded among equivalent type spaces via *partially correlating devices* — devices that send correlated signals to players in a way that preserves each player’s belief about the others’ types; second, since the belief-invariant Bayesian solution is itself implemented by exactly such devices, it is invariant across equivalent type spaces. So the architecture is the same as ours and Liu’s — identify the class of devices/expansions that the solution concept tolerates, show that moving between hierarchy-equivalent spaces is exactly an operation in that class — with the device class (“partially correlating”) playing the role our proof assigns to hierarchy-preserving expansions.

There's a companion paper, Tang (2015, GEB 91, 36–44), on the rationalizability side: it formalizes interim partially correlated rationalizability (IPCR), which permits correlations that depend on opponents' types but not on the state, and shows that Δ -hierarchies are necessary and sufficient to identify IPCR, as a direct extension of Ely and Pęski's main theorem from two-player to n -player games; the type-correlated conjectures there again take the form of partially correlating devices, linking back to the JME paper. The pair of papers slots into the map as follows: Dekel, Fudenberg, and Morris (2007) show Mertens–Zamir hierarchies suffice for interim *correlated* rationalizability, EP show they fail for interim *independent* rationalizability and that Δ -hierarchies are the right invariant there, and Tang shows Δ -hierarchies are also the right invariant for IPCR and for the belief-invariant Bayesian solution.

The correction. Last turn I said Tang shows the belief-invariant Bayesian solution “is determined by hierarchies”; the accurate statement is that it's determined by *Ely–Pęski Δ -hierarchies*, which are strictly richer than Mertens–Zamir hierarchies. That distinction is the interesting part. With a common prior, Liu (2015) gets invariance of belief invariant BCE in ordinary MZ hierarchies; without the common prior discipline, the Bayesian-solution family needs the finer EP invariant. Our Theorem 1 sits squarely on the Liu branch — common prior, MZ-style hierarchies, upgraded from players to groups — so in the related-literature paragraph Tang should be cited as the non-common-prior cousin with a *different* sufficient statistic, not as a hierarchy-invariance result parallel to ours.

One speculative but suggestive connection to our shelved open question: the reason Δ -hierarchies arise in Tang's setting is that his devices only preserve beliefs about others' *types*, allowing residual correlation that ordinary hierarchies can't see — structurally similar to how our disjoint-groups invariance permits a_j to correlate with intersecting groups' information in ways invisible to group hierarchies. If the weak-GBI version of Theorem 1 fails (the purification question), the natural repair would be an EP-style enrichment of *group* hierarchies as the correct sufficient statistic for the weak notion — a “group Δ -hierarchy.” That would make the strong/weak pair in our paper a precise group-level analogue of the DFM/EP dichotomy, which is a nice story if it pans out. Worth keeping in the ideas file for after the current draft stabilizes.

Ben & Songzi

Please add the minimality result to a new subsection immediately after the subsection containing Theorem 1.

Claude

The minimality result is in place as new subsection 3.4, “Minimality of Group Belief Invariant BCE,” between the sufficiency subsection and the additional-properties subsection (now 3.5). Files below, 20 pages, compiling cleanly.

The subsection contains, in order:

Definitions. Bayes Nash equilibrium for group type spaces (the obvious extension: expected utility $U_i(b; \mathbb{I}, G)$ with strategies on singleton types, and the ex-ante inequalities); solution concepts as maps from group type spaces to subsets of $\Delta(A \times \Theta)$; the examples S^{BNE} and S^{GBI} ($\mathbb{I} \mapsto S_{\mathbb{I}}(G)$); group belief invariance of a solution concept ($\pi_{\mathbb{I}} = \pi_{\mathbb{I}'} \Rightarrow S(\mathbb{I}) = S(\mathbb{I}')$); and refinement. A remark records your observation that S^{BNE} is *not* invariant, via the payoff-irrelevant public coin,

which leaves every group’s hierarchy — hence π — unchanged but enlarges the BNE set whenever correlation helps.

Theorem 2. S^{GBI} is group belief invariant, is refined by S^{BNE} , refines every group belief invariant solution concept that S^{BNE} refines, and satisfies the union formula $S^{GBI}(\mathbb{I}) = \bigcup \{S^{BNE}(\mathbb{I}') : \pi_{\mathbb{I}'} = \pi_{\mathbb{I}}\}$ — followed by the one-sentence gloss (minimal invariant concept containing the BNE predictions) and the attribution sentence framing it as the group analogue of Liu (2015).

Proof, in three steps. Step 1: BNE outcomes are GBI BCE — obedience via the one-line deviation argument (LHS = $b_i(a_i \mid \tau)D$ with $D < 0$ contradicting equilibrium), invariance because the conditional law of a_j given (t, θ) is $\prod_{i \in J} b_i(\cdot \mid t_{\{i\}})$, a function of t_j by degeneracy. Step 2: the revelation-with-hierarchy-preservation construction — append recommendations, $T'_j = T_j \times A_j$; verify $\mathbb{I}'$ is a group type space (subvector plus degeneracy); prove $h'_j(t_j, a_j) = h_j(t_j)$ by simultaneous induction on belief orders, using at each order exactly the invariance property that conditioning on a_j beyond t_j leaves the conditional law of any function of $(t_{\mathcal{J} \setminus \{j\}}, \theta)$ unchanged; conclude $\pi_{\mathbb{I}'} = \pi_{\mathbb{I}}$, and check the obedient strategies form a BNE inducing ρ . Step 3 assembles both inclusions of the union formula and the minimality claim.

Two notes. Theorem 1’s standing positive-probability assumption (“for the remainder of this section”) covers the new subsection, and the construction in Step 2 invokes the zero-probability-deletion normalization explicitly. And since Theorem 2’s proof leans on the strong (subgroup-measurability) form of invariance precisely at the hierarchy-preservation induction, this subsection is where the definitional choice from earlier now visibly pays off — if you ever want a remark to that effect in the text (that under the disjoint-groups variant, Step 2 is exactly the open purification question), it would fit naturally right after the theorem’s gloss.

[Files delivered: *groupbeliefs.pdf*, *groupbeliefs.tex*]

Ben & Songzi

Could you timestamp the drafts in their titles?

Claude

Done — the current draft is now delivered as *groupbeliefs_9-17-2026.tex* and *groupbeliefs_9-17-2026.pdf*, following the same pattern as your *seconddraft_9-15-2026_bab.tex*, and I’ll carry the date stamp forward on every delivery from here on (updating it when the calendar turns). The title page itself already carries the compilation date via `\date{\today}`, so the document and filename agree.

Two small options if you want the convention adjusted: I can append an initials suffix as in your BDH file (e.g., `_bd` or `_bab`), or switch to sortable ISO dates (*groupbeliefs_2026-09-17.tex*), which keeps drafts in chronological order in a directory listing. Just say which pattern you and Songzi use.

[Files delivered: *groupbeliefs_9-17-2026.pdf*, *groupbeliefs_9-17-2026.tex*]

Ben & Songzi

Please update the transcript. Also note that even though this is Ben’s account, you have been conversing with Ben and Songzi.

[This updated transcript was produced in response to the preceding message.]