Notes on Group! Agency[1]

Brian Epstein
Tufts University, Medford

Michael D. Ryall
University of Toronto

April 2, 2021

[1]

1 Overview

This version begins with a model of individual agency in Section 3, then moves on to groups and group agency in the remaining sections. We are aiming for a formal framework that is fairly general, thereby allowing for a substantial degree of flexibility in the sorts of phenomena it can represent. The formalism for groups builds on the individual setup.

2 Notational conventions

2.0.1 General

Capital letters (G, N, etc.) refer to sets. Small Arabic and Greek letters refer variously to elements of sets (e.g., $i \in N$) and functions (e.g., $\sigma : N \to \mathcal{N}$). Terms are *italicized* at the point of definition. A *profile* is a placeholder for a list of elements. We denote these in boldface: e.g., $\mathbf{x}$ where $\mathbf{x} \equiv (x_1, \ldots, x_n)$. The "$\equiv$" symbol indicates the definition of a mathematical object. If X is a set, then 2^X is the notation denoting the set of all subsets of X. Calligraphic letters refer to sets of sets (e.g. $\mathcal{X} \equiv 2^X$). Curly parentheses indicate sets, typically in defining them (e.g. $X \equiv \{x | x \text{ is an even integer}\}$). The notation "$|\cdot|$" indicates set cardinality (e.g., if $X \equiv \{a, b, c\}$, then $|X| = 3$). If X is a set and $Y \subset X$, then $X \setminus Y$ is the set X minus Y; i.e., the set of elements of X that remain when the elements of Y are removed. All sets are assumed to be finite unless otherwise indicated.

2.0.2 Specific

The following table elaborates all the mathematical objects used in the paper.

3 Individual Agency

Begin with a *population of individuals*, indexed by the set $N \equiv \{0, \ldots, n\}$ with typical element $i \in N$ and $\mathcal{N} \equiv 2^N$. For now, we focus on an individual actor. Later, we consider groups. The evolution of the world through time is driven by the actions of individuals as well as of the onset of natural phenomena. We account for natural phenomena as the "actions" of Nature which we assign to population index 0.

We break this section into two subsections. The first develops the mathematical machinery to discuss and analyze actual and potential states of the world at a moment in time. We refer to this as the *synchronic* perspective. With these details in place, we then extend the framework to the dynamic case, in which the world evolves through time. We refer to this as the *diachronic* perspective.

3.1 Synchronic Setup

3.1.1 States

A *state*, denoted s, is a snapshot of the world at a moment in time. States elaborate the status *of all features of the world* in that moment. This includes the relevant "mind-independent" features of a particular world as well as the "mind-dependent" features of the individuals acting in that world. Clearly, it would require an uncountably infinite number of states to elaborate everything about the world in a given moment, much less all the potential features that could be actualized in that moment. However, our discussion will always focus upon a finite set of actors who are typically concerned with a particular set of issues. Therefore, we sidestep some mathematical complexities by limiting our attention to the relevant features by assuming that the potential number of states required to describe the features of interest at any particular moment are finite.

With this in mind, let S^0 denote the (finite) set of all possible states of the world. The "0" superscript indicates that this corresponds to Nature's "perception" of reality; i.e., $s \in S^0$ is a description of the real world as it could actually exist. Individual superscripts, e.g., S^i, $i \neq 0$, indicate individual i's (typically, limited) awareness of reality. Specifically, we assume $\mathcal{S} \equiv \{S^\emptyset, S^0, S^1, \ldots, S^n\}$ along with $\succeq$, a partial order on $\mathcal{S}$, is a complete lattice in which S^0 is a maximum (the richest expression of reality) and $S^\emptyset$ is a minimum (the poorest expression), where $S^\emptyset \equiv \{\emptyset\}$ is defined as the state space consisting of a single element (i.e., in which nothing about the world is distinguished).[1] Let $\Sigma \equiv \bigcup_{i \in N} S^i$ denote the union of the individual spaces.

Conceptually, for an actual individual i, $s^i \in S^i$ includes all the features of reality that individual i can bring to mind and relate to in the moment at hand. Thus, $S^i \succeq S^j$ means that individual i is able to distinguish at least as much about the world as individual j in that moment. Because $\succeq$ is a partial order, not all state spaces are comparable; i.e., individuals i and j may be aware of different things in a given moment.

[1] Here, we adapt the interactive unawareness approach developed by **?**

We wish to keep track of how the different state spaces relate to reality (S^0) and, when possible, to each other. Therefore, define the surjective *projection* $r^{i\to j} : S^i \to S^j$, which is only defined if $S^i \succeq S^j$. Then, $s^j = r^{i\to j}(s^i)$ is the impoverished version of reality j perceives relative to the awareness of i. By assumption, for all $i \neq 0$, $S^0 \succeq S^i$. Assume the projections are commutative: if $S^i \succeq S^j \succeq S^k$, then $r^{i\to k} = r^{j\to k} \circ r^{i\to j}$.

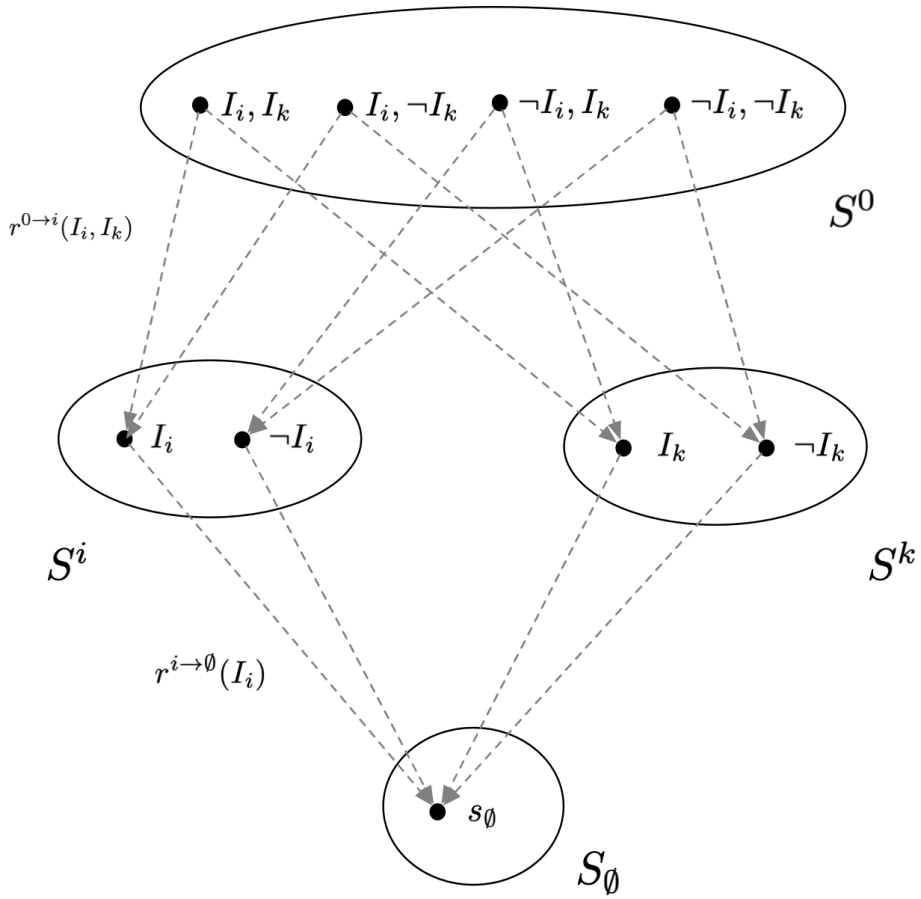

Figure 1: Awareness of Irene and Ken

To see a simple example of the setup, consider a situation in which Irene (i) and Ken (k) form an intention of whether or not to take a walk together. Let I_i and $\neg I_i$ indicate that Irene intends or does not intend, respectively, to take a walk with Ken and, similarly, for Ken. Then, in this simple world, there are four states that can be true: $(I_i, I_k), (I_i, \neg I_k), (\neg I_i, I_k)$ and $(\neg I_i, \neg I_k)$. In Figure 1, we see the individual spaces represent a world in which Irene and Ken are only aware of their own intentions. The projections are shown as dashed lines, with $r^{0\to i}(I_i, I_k)$ and $r^{i\to\emptyset}(I_i)$ specifically labelled. Thus, if (I_i, I_k) is the true state of the world, then Irene is only aware of her

intention and, similarly, Ken is only aware of his intention. As required, the figure includes the state of complete unawareness, $S_\emptyset$. Note that, while $S_t^0 \succeq S_t^i \succeq S_\emptyset$ and $S_t^0 \succeq S_t^k \succeq S_\emptyset$, S_t^i and S_t^k are neither richer nor poorer than the other. Here, Σ is the set containing all the states from all the state spaces.

3.1.2 Synchronic events

The term 'event' is used differently in philosophy than it is in probability theory. Since we are writing to audiences familiar with one or the other, it is important to clarify this difference. In probability theory, 'event' is used similarly to the term 'property' in philosophy, where properties are understood intensionally. Philosophers typically use 'event' to mean a spatiotemporal particular extended over time. We refer to events associated with states at a moment in time (the game theory useage) as *synchronic events*, and those associated with states unfolding through time (the philosophy usage) as *diachronic events*. Below, we define the former. We wait to define the latter until Section 3.2.

In probability theory, events are subsets of state spaces. For example, the event "Mike intends to get a cup of coffee includes *all* states in which getting a cup of coffee is the intention of Mike. In philosophical terminology, this is equivalent to the property *being in a state in which Mike intends to get a cup of coffee,* where the intension of the property is all the states of the world in which the world exemplifies that property. Because each individual is associated with a state space that elaborates states according to the features of the world of which that individual could be aware in a given moment, the events of which he or she could be aware are subsets of that space. For example, $E = \{s \in S^i | s \Rightarrow \text{Mike has coffee}\}$ is the event, which i is aware of as a possibility, that Mike has a cup of coffee (i.e., all the states in which this obtains according to i).

Because individual state spaces may be related to one another and, in any case, are all related to reality fully elaborated (S^0), it will be helpful to consider the events that can be described in one individual space to those implied in spaces that are at least equally as rich. To set this up, let $g : \mathcal{S} \to 2^{\mathcal{S}}$ where $g(S^j) \equiv \{S^i \in \mathcal{S} | S^i \succeq S^j\}$ identifies the set of state spaces that are at least as rich as S^j. For a state-space event $B \subseteq S^j$, let $B^\uparrow = \bigcup_{S^i \in g(S^j)} \left(r^{i \to j}\right)^{-1}(B)$ be the extension of B to include all states in other individuals' state spaces that provide elaborations of reality that are at least as rich as S^j. Then, $E \subseteq \Sigma$ is a *synchronic event* if it is of the form $B^\uparrow$ for some $B \subseteq S^i \in \mathcal{S}$. We refer to the state space event, B, as the *basis* of the extended event $E = B^\uparrow$ and to S^j as the

base-space of E. By this definition, not every subset of Σ is a synchronic event. If $B \subseteq S^j$, define the negation of the synchronic event $B^{\uparrow}$, denoted $\neg B^{\uparrow}$, as $(S^j \setminus B)^{\uparrow}$, typically a proper subset of $\Sigma \setminus B^{\uparrow}$.

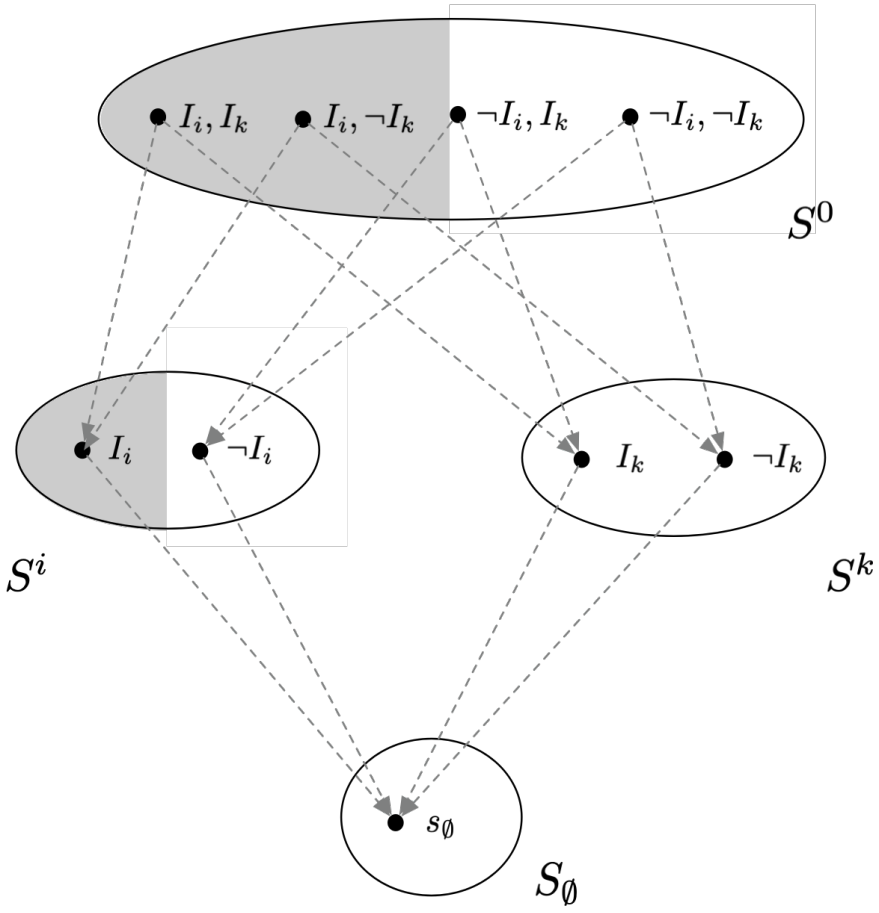

Figure 2: The awareness structure event "Irene intends to walk with Ken"

Returning to our example with Irene and Ken, consider the event "Irene intends to take a walk with Ken" in S^i. Let $B = \{I_i\} \subseteq S^i$ be this state-space event. Then, B is the basis of $B^{\uparrow} = \{(I_i, I_k), (I_i, \neg I_k), I_i\}$ and S^i is the base-space. Notice that $\neg B^{\uparrow} = \{(\neg I_i, I_k), (\neg I_i, \neg I_k), \neg I_i\}$. Thus, $B^{\uparrow} \cup \neg B^{\uparrow}$ is a strict subset of Σ. We also see that the states in an individual state space represent a coarsening of reality (formally, a partition of S^0) due to the fact that the projections are surjective functions from more refined spaces to coarser spaces.

++STOPPED HERE++++++++++++++++++

3.1.3 Mental attitudes

In what follows, we develop an awareness-belief-desire-intention model of mental attitudes. The essential aim of this formalism is to take seriously the cognitive constraints we face as finite, material beings. In particular, we proceed from the uncontroversial claim that, at any given moment, an individual can only attend to some finite number of conscious concerns. We say that an individual is *aware* of the matters toward which his or her attention is directed. Under constrained awareness, intentions take on an important role that is distinct from beliefs and desires.

The idea is as follows. To the extent some share of the mind's resources are occupied in solving a problem (e.g., deciding what kind of car to buy), those resources are not available for other conscious operations, such as solving other problems, constructing a feasible plan by which to acquire a car, or actualizing that plan by driving to the car dealer and making the transaction. We conjecture that an individual's finite stock of cognitive resources almost always acts as a hard constraint on his or her decision- and act-making capability. In our model, intentions serve as the pivot from goal assessment to goal acquisition. The formation of an intention moves an individual from a state in which an individual is reckoning what to do to a new state in which the individual has decided what to do and in which he or she has at least some sense of how to proceed – i.e., a *plan*.[2] Thus, forming an intention frees up the mental resources required to determine which goal to pursue and how to pursue it. When events arise consistent with the plan, the individual can act accordingly – without engaging the mental machinery required to reassess goals and plans. Because deciding to focus attention on some new problem can, itself, be an intentional goal, one's awareness is dynamic and, to some extent, influenced by one's own intentions. As we will see, there are also social implications as individuals become aware of the intentions of others.

Beliefs and desires will operate in a familiar way. The distinction here is that they are restricted to those matters about which an individual is aware. As we show below, because beliefs cannot account for awareness and because intentions shift awareness, a belief-desire model cannot do the work of an awareness-belief-desire-intention model.

Awareness There are two conditions that must be met for an individual to be aware of some information. First, the information must be accessible to oneself for active consideration. The

[2] A more elaborate treatment might well separate each step by an act of intention: first, the move from goal assessment to plan selection; then, from plan selection to plan implementation. For now, we bundle these steps into one.

sources of accessible information are contemporaneous sense data, active imagination, and knowledge – essentially, anything an individual can call to mind. Second, the information must be actively brought to mind. For example, an airline pilot may be able to call to mind how to navigate a jetliner but not a container ship. That same pilot may not be aware of how to navigate a jetliner while driving his or her car down the freeway. We cannot bring to mind things we do not know or cannot imagine. Of the things we know or can imagine, we are constrained in the number to which we can actively attend.

Unawareness has long been a tricky problem for decision theorists. A decision maker can only choose between acts of which he or she is aware which, typically, does not include all the truly feasible acts at that moment. Moreover, the decision problem is further compounded by unawareness of future possibilities associated with one's acts. It is easy enough to represent a static decision problem which is constrained by the decision maker's awareness of possible acts by simply defining the "feasible" acts as those corresponding to his or her awareness. The problem is how to model what happens in a dynamic setting in which the decision maker suddenly faces an unexpected consequence. For example, in a standard Bayesian decision problem, unawareness of certain consequences can be modeled as zero-probability states according to the decision maker's subjective beliefs. However, such decision makers will be confounded should a subjectively impossible state occur. Added to this is the problem of representing decision makers of differing awareness when decision problems are interactive.

Dekel et al. (1998) demonstrate that standard state-space approaches cannot model unawareness. Schipper (2015) surveys various alternatives to modeling unawareness, including approaches from AI, logic, and game theory. We adopt a version of the framework used in Bryan et al. (2021) which itself builds on previous work developed in Heifetz et al. (2006), Heifetz et al. (2008), and Heifetz et al. (2013). This approach solves the problems mentioned above by creating multiple state spaces, each one associated with the awareness of a particular individual. This allows different agents to have different perceptions of the the true state of the world as well as the future states that might obtain in the future.

When t is a future state, an "awareness of awareness" complication arises. That is, does $r_t^{0\to i}(s_t^0)$ represent the all the features of s_t^0 of which i would be aware in an objective sense? Or, does it represent the all the features of the world that i, in the present period, subjectively imagines she could be aware of in period t? For example, Sam may believe that in the next moment, she

can become aware of the temperature outside by checking the internet. Yet, suppose Nature is going to crash her internet service with certainty. Then, Sam's presumption of future awareness is not correct. The answer, according to our logic, is that i's awareness of her future potential for awareness is a mental attitude that is also embedded within the *present* state. Thus, if need be, we can extend $r_t^{0\to i}$ so that, for any $T \geq w \geq t$, $s_w^i = r_t^{0\to i}(s_t^0, s_w^0)$ represents i's awareness in state s_t^0 about the future state s_w^0.

In general, $r^{k\to i}$ is not defined unless $S_t^k \succeq S_t^i$. Given the information encoded in a state, individuals may also be aware of what they know, what they believe, what they intend, and so on. Importantly, awareness may extend to the mental states of others.

For example, suppose the only information of interest in the first period is the outcome of a single die roll. Then, $S_1^0 = \{1, 2, 3, 4, 5, 6\}$. Further, suppose that individual i is told "red" if 1 or 2, "blue" if 3 or 4, and "green" if 5 or 6. Then, $S_1^i = \{red, blue, green\}$ and $r_1^{0\to i}(1) = r_1^i(2) = red$, etc. Notice that the states in S_1^i correspond to synchronous events in S_1^0.

As was explained earlier, in a standard state space, synchronous events are subsets of states that, typically, correspond to some true aspect of the world – e.g., "the price of gold hit a new high on January 23" is the synchronous event containing all states of the world in which that description is true. Through the projections, synchronous events in S_t^0 are associated with all the other state spaces in the lattice. Given the conditions on the r_t^is, every event in Nature's state space maps to events in all awareness spaces. The S_t^0 event "Irene intends to take a walk with Ken" extended to the awareness structure is illustrated by the shaded area in Figure 2. Irene knows whether she intends to walk with Ken, but Ken is unaware of that intention. If $E_t \in \Sigma_t^0$ is a sycnronous event in S_t^0, then we denote by $\gamma_t(E_t) \subseteq \Gamma_t$ the states that correspond to E_t in the awareness lattice. Then, for example, the event within Irene's awareness structure that corresponds to Nature's event E_t is $\gamma_t(E_t) \cap S_t^i$.

++

Beliefs Beginning with beliefs, let $\Delta(H)$ denote the set of all probability distributions on the set of histories. Then, $\mu_i : S \to \Delta(H)$ is a function that maps from states to individual i's beliefs on histories H. We write μ_i^s to indicate i's subjective probability distribution on H at state s. This distribution induces a distribution on history events, $\mathcal{H} \equiv 2^H$. Note that each μ_i^s induces a probability distribution on S. For example, the probabilities of the elements of Z (terminal nodes) are equal to the probabilities of the complete histories they terminate. The probability of some arbitrary state s_t is equal to the sum of the probabilities of the complete histories running through it, and so on. Since all of this is implied by μ_i, we will slightly abuse notation and write, e.g., $\mu_i^s(Z) = \mu_i^s(H)$, even though $Z \in \mathcal{S}$ while $H \in \mathcal{H}$.

It is important to note that the existence of more than one element in S_0 means that individuals may be uncertain about which tree is the objective one and, hence, the true history they have experienced. If so, they will be uncertain about which state they are in. In addition, there will be uncertainty about how the future unfolds. At the moment, we have the objective world starting at s_0^* and unfolding in accordance with ω and the sequence of everyone's act choices. Since acts are free choices by individuals, it is possible they are selected randomly ("now, I will decide what to do by flipping a coin"). This includes acts of Nature. All of individual i's speculation with respect to the history, state and unfolding of events is summarized by μ_i.

Like in the case of incomplete information, we proceed by introducing probability distributions on state-spaces. For any state space $S \in \mathcal{S}$, let $\Delta(S)$ be the set of probability distributions on S. Even though we consider probability distributions on each space $S \in \mathcal{S}$, we can talk about probability of events that, as we just have seen, are defined across spaces. To extend probabilities to events of our lattice structure, let S_μ denote the space on which μ is a probability measure. Whenever for some event $E \in \Sigma$ we have $S_\mu \succeq S(E)$ (i.e., the event E can be expressed in space S_μ) then we abuse notation slightly and write

$$\mu(E) = \mu(E \cap S_\mu).$$

If $S(E) \not\preceq S_\mu$ (i.e, the event E is not expressible in the space S_μ because either S_μ is strictly poorer than $S(E)$ or S_μ and $S(E)$ are incomparable), then we leave $\mu(E)$ undefined.

To model an agent's awareness of events and beliefs over events and awareness and beliefs of

other groups, we introduce type mappings. Given the preceding paragraph, we see how the belief of an agent at state $\omega \in S$ may be described by a probability distribution over states in a less expressive space S' (i.e., $S \succeq S'$). This would represent an agent who is unaware of the events that can be expressed in S but not in S'. These events are "out of mind" for him in the sense that he does not even form beliefs about them at ω: his beliefs are restricted to a space that cannot express these events.

More formally, for every agent $i \in N$ there is a *type mapping* $t_i : \Omega \longrightarrow \bigcup_{S \in \mathcal{S}} \Delta(S)$. That is, the type mapping of agent $i \in N$ assigns to each state $\omega \in \Omega$ of the lattice a probability distribution over some space. Now a state does not only specify which events affecting value creation may obtain, and which beliefs agents hold over those events, but also which events agents are aware of. Recall that S_μ is the space on which μ is a probability distribution. Since $t_i(\omega)$ now refers to agent i's probabilistic belief in state ω, we can write $S_{t_i(\omega)}$ as the space on which $t_i(\omega)$ is a probability distribution. $S_{t_i(\omega)}$ represents the *awareness level* of agent i at state ω. This terminology is intuitive because at ω agent i forms beliefs about *all* events in $S_{t_i(\omega)}$.

For a type mapping to make sense, certain properties must be satisfied. The most immediate one is *Confinement:* if $\omega \in S'$ then $t_i(\omega) \in \triangle(S)$ for some $S \preceq S'$. That is, the space over which agent i has beliefs in ω is weakly less expressive than the space contains that ω. Obviously, a state in a less expressive space cannot describe beliefs over events that can only be expressed in a richer space. We also impose Introspection, which played a role in our prior discussion of incomplete information: every agent at every state is certain of her beliefs at that state. In AppendixXX, we discuss additional properties that guarantee the consistent fit of beliefs and awareness across different state-spaces and rule out mistakes in information processing.

It might be helpful to illustrate type mappings with an example. FigureXX depicts the same lattice of spaces as in FiguresXX and XX. In addition, we depict the type mappings for three different groups. At any state in the upmost space S_{pq}, the blue agent is aware of p but unaware of q. Moreover, she is certain whether or not p depending on whether or not p obtains. This is modeled by her type mapping that assigns probability 1 to state p in every state where p obtains and probability 1 to state $\neg p$ in every state where $\neg p$ obtains. (The blue circles represent the support of her probability distribution that must assign probability 1 to the unique state in the support.) An analogous interpretation applies to the red agent except that she is an expert in q. In contrast, the green agent is aware of both p and q but knows nothing with certainty, modeled

by her probabilistic beliefs in the upmost space that assigns equal probability to each state in it.[3]

Unawareness structures allow us to model an agent's awareness and beliefs about another agent's awareness and beliefs, beliefs about that, and so on. This is because, as in the incomplete information case, beliefs are over states and states also describe the awareness and beliefs of groups. Return to FigureXX. At state pq the green agent assigns probability 1 that the blue group is aware of p but unaware of q. Moreover, he assigns probability 1 to the blue agent believing with probability 1 that the red group is unaware of p.[4]

Desires For all $i \in N$, define the state-dependent *desire relation* such that, for all $s \in S$, $D_i^s \subset P \times P$ where, $(p', p'') \in D_i^s$ means that individual i in state s desires the path p'' at least as much as the path p'. Having described the mathematical structure of desires, we use the more intuitive notation $p' \preceq_i^s p''$, which is defined to mean $(p', p'') \in D_i^s$. We use $\prec_i^s$ and $\approx_i^s$ to indicate strict preference and indifference, respectively.

Why make preferences over paths? Because we assume individuals care about how they get to an end as well as the end itself. To take a canonical example, a homeowner may have a renovated kitchen in mind as the desired end. However, even if the kitchen specs are provided in extensive detail (so the owner knows exactly what the end will be), there may be many contractors who can deliver it. In this case, assuming there are several contractors from which to choose, each of which identify with a different path with states encoding costs at each step of the way and the final quality of the work, the owner's choice will be based upon the path (costs) as well as the final state (quality). Similarly, an individual sensitive to the time value of money will prefer shorter paths to longer ones, other things equal. Or, individuals may value portions of the paths themselves. For example, even though a student drops out of school (thereby, not completing the degree), he or she may nevertheless value the portion of the education that was completed. Our approach allows for special cases in which all these details are elaborated as primitives of the situation. For our discussion, we simply assume preferences are over paths.

[3]The example is taken from Schipper (2016) who shows how a generalist (i.e., the green agent) emerges as an entrepreneur and forms a firm made of specialists (i.e., the blue or red agents) in a knowledge-belief and awareness-based theory of the firm using strategic network formations games under incomplete information and unawareness.

[4]We note, it has been shown that under appropriate assumptions on spaces $S \in \mathcal{S}$ and the type mapping, unawareness structures are rich enough to model any higher order beliefs of agents (see the working paper version of Heifetz et al. (2013)).

Intentions Finally, define the state-contingent *intention* for individual i as a function $\gamma_i : S \to \mathcal{S}$, where $\gamma_i(s) = E$ means that in state s individual i intends event E. We assume that individuals have desires and beliefs in all states, but not necessarily intentions. The idea here is that, e.g., in some states Mike intends the end "Mike has a cup of coffee" and in others, Mike has yet to form intentions. We adopt the convention that $\gamma_i(s) = \emptyset$ means that s is a state in which individual i has not formed an intention. We highlight that states may be differentiated only by changes in mental attitudes. For example, it may be that the only change from s_t to s_{t+1} is $\gamma_i^{s_t} = \emptyset$ to $\gamma_i^{s_{t+1}} = E$. This suggests that the interval between time periods may be very short (measured in milliseconds).

This raises the question of how an individual moves from being in a state without an intention to one in which the intention is formed. Here, we can require an act of commitment to cement the intention. That is, if s_t is a state in which i does not have an intention, then the set of feasible acts, $A_i^{s_t}$, can include an *act to form the intention* to "get a cup of coffee," which would then take him to a state s_{t+1} in which $\gamma_i^{s_{t+1}} = X$ where X contains all the states consistent with i having a cup of coffee.

For all $i \in N$, individual i's *mental attitudes* are summarized by a triple denoted $\theta_i \equiv (\mu_i, D_i, \gamma_i)$.[5] A *profile of mental features* for all the individuals is given by the profile $\theta \equiv (\theta_1, \ldots, \theta_n)$. Given our conventions, we can write $\theta_i(s)$ and θ^s without ambiguity.

3.2 Diachronic setup

Assume time is discreet and limit attention to some finite number of periods, T. *Nature's state space at time t*, denoted $S_t^0 \subset S^0$, with typical element $s_t^0 \in S_t^0$, contains all the states that could possibly be actualized at t.[6]

3.2.1 Acts and actions

The sequence of states actualized over the period of analysis is effected by the acts of the individuals in the population in conjunction with acts of Nature (i.e., all the causes that, in conjunction with the acts of the individuals, determine the actualization of a particular state from an immediately preceding, previously actualized state). Let A be the set of all possible acts that can be chosen by

[5] In setting up mental features in this way, we are following a version of the familiar "type-space" approach used in game theory (See Harsanyi, 1967; Mertens and Zamir, 1985).

[6] This setup can be generalized to include uncountably infinite state spaces. Limiting attention to finite sets allows us to sidestep some mathematical complexities which would add little to our analysis.

some individual in some state. For each individual $i \in N$, $A^i(s_t^0)$ indicates the set of *feasible acts available to individual i in state* s_t^0 with typical element $a_t^i \in A^i(s_t^0)$.[7]

We adopt the convention that $A^i(s_t) = \emptyset$ indicates that individual i has no available acts in state s_t. An *act profile* is a list of acts, one for each individual, denoted $\mathbf{a}_t \equiv (a_t^0, a_t^1, \ldots, a_t^n)$. Recall, Nature is "Individual 0" so that a_t^0 summarizes all the developments that, in conjunction with the individuals' acts, determine which state is actualized following s_t. The set of *all act profiles at state* s_t is $\mathbf{A}(s_t) \equiv \times_{i=0}^n A^i(s_t)$; the set of *all possible act profiles at time t* is $\mathbf{A}_t \equiv \cup_{s_t \in S_t} \mathbf{A}(s_t)$; and the set of *all possible act profiles* is $\mathbf{A} \equiv \cup_{s \in S} \mathbf{A}(s)$.

3.2.2 Dynamics

As indicated above, the act profiles summarize all the contingencies required to actualize one state from the next. To formalize this, let $\omega : \mathbf{A} \times S^0 \to S^0$ be the *state-contingent actualization function*, where $\omega(\mathbf{a}_t, s_t^0) = s_{t+1}^0$ indicates that if the act profile at state $s_t^0 \in S_t^0$ is $\mathbf{a}_t \in \mathbf{A}(s_t^0)$, then the next state actualized is s_{t+1}^0. Assume that, for all t, ω is bijective from $\mathbf{A}_t \times S_t$ to S_t^0. In other words, each feasible act profile in a given state at time t leads to a unique state in period $t+1$ and each state in period $t+1$ can be traced back to a single predecessor state in period t by a unique act profile that links the two. Thus, the inverse ω^{-1} exists, where $\omega^{-1}(s_{t+1}^0) = (\mathbf{a}_t, s_t^0)$ indicates that s_{t+1}^0 is actualized when $\mathbf{a}_t$ is enacted in s_t^0.

Suppose, for example, that two distinct sequences of acts could lead to an identical footprint in the snow. In that case, we consider there to be two states in which that identical footprint exists, each associated with one of the sequences of acts that lead to it.

The world begins at state s_0^0. To allow for uncertainty or partial knowledge with respect to various aspects of the world at the beginning of time, we assume Nature's acts entirely determine s_1^0. That is, $\mathbf{a}_0 = (a_0^0, \emptyset, \ldots, \emptyset)$, where a_0^0 represents all the actualized historic factors that lead individuals to their first decision state, $s_1^0 = \omega(\mathbf{a}_0 | s_0^0)$. Uncertainty with respect to the state of the world in $t = 1$ (e.g., about the intentions or other individuals) is, thus, formalized as uncertainty about "Nature's act" $a_0^0 \in A^0(s_0)$ prior to the first decision period.

We define the *history at state* s_t^0 as a profile of states that starts at s_0^0 and ends at s_t^0, de-

[7] Notice that we use a capital letter to indicate that A^i is a set-valued function: $A^i : S^0 \to 2^A$. Also note that feasible acts for individual $i \neq 0$ are determined by reality (states in S^0), not by i's awareness of reality (states in S^i). Because we consider the intentional formation of some mental attitudes as choices available to individuals, we use the term "act" to describe the choices available to someone in a broad way. We think of "action" as describing the narrower category of act associated with physical movement.

noted $\mathbf{h}(s_t^0) = (s_0^0 \ldots, s_t^0)$. A history $\mathbf{h}(s_t^0)$ is *feasible* if there exists a sequence of action profiles $\mathbf{a}_0, \ldots, \mathbf{a}_{t-1}$ such that $s_1^0 = \omega(\mathbf{a}_0|s_0^0), \ldots, s_t^0 = \omega(\mathbf{a}_{t-1}, s_{t-1}^0)$. Clearly, feasible histories are the only ones that can be actualized according to objective reality. This distinction allows for situations in which individuals subjectively consider infeasible histories to be possible. For example, individual i may believe that act $a_t^i \in A^i(s_t^0)$ is consistent with the actualization of s_{t+1} even though a_t^i is not in the profile $\mathbf{a}_t$ that leads from s_t to s_{t+1}.

The set of all *histories at time t* is $\mathbf{H}_t$. The set of all histories is $\mathbf{H}_T$ and the set of all subsets of histories is $\mathcal{H}_T$. According to our notational convention, adding a "0" superscript to these objects refers to their feasible counterparts; e.g., the set of all *feasible histories at time t* is $\mathbf{H}_t^0$. An arbitrary *history at time t* is denoted $\mathbf{h}_t \in \mathbf{H}_t$, where we start with the *null history* $\mathbf{h}_0^0 = (s_0^0)$ at the beginning of time (so, $\mathbf{H}_0^0 = \{\mathbf{h}_0^0\}$ and $S_0^0 = \{s_0^0\}$). Because there is a single root node and ω is a bijection, the set of paths in $\mathbf{H}_T$ form a tree. Thus, S^0 can be partitioned according to subsets of states corresponding to time periods: $S^0 = S_0^0 \cup \cdots \cup S_T^0$ and $S_0^0 \cap \cdots \cap S_T^0 = \emptyset$. Note also that each S_t implies a partition of $\mathbf{H}_T$ according to the sets of paths intersecting the states in S_t.

3.2.3 Diachronic events

A *diachronic event* is a subset $E \in 2^{\mathbf{H}_T}$; i.e., a subset of paths in the tree associated with $\mathbf{H}_T$.[8] Let $\mathcal{E} \equiv 2^{\mathbf{H}_T}$ be the set of all diachronic events. Given the preceding discussion, every synchronic event $\sigma_t \in \Sigma_t$ is associated with a unique event $E \in \mathcal{E}$.

To see how we use states and understand how these objects work, consider the canonical example of rolling a six-sided die. We use functions on S^0 to "extract" information from the states. Here, for example, we can let $d(s_t^0)$ indicate the outcome of a die roll in state s_t^0: for all $s_t^0 \in S_t^0$, $d(s_t^0) \in \{1, 2, 3, 4, 5, 6\}$; i.e., d maps from each state in S_t^0 to a number between 1 and 6, indicating the side of the die that landed up in that state (where s_t^0 includes *all* features of the world besides how the die landed). Now, the synchronic event "the die roll is even" is described by $\sigma_t \in \Sigma_t$ such that $\Sigma_t \equiv \{s_t^0 \in S_t^0 | d(s_t^0) = 2, 4 \text{ or } 6\}$. Alternatively, suppose $T = 2$. Then, the diachronic event, "snake-eyes were rolled" is described by $E \in \mathcal{E}$ such that $E \equiv \{(s_0^0, s_1^0, s_2^0) \in \mathbf{H}_T | d(s_1^0) = d(s_2^0) = 1\}$.

[8] Note that diachronic events are subsets of whole paths from s_0^0 to subsets of states in S_T^0. Therefore, they do not have time subscripts.

3.3 Consistency conditions

Having structured the objects of interest, we now explore various conditions required to impose the regularities between the various mental attitudes and between those attitudes and the external world that are appropriate to a rational human being.

Reality Alignment Beginning with the latter, our setup allows individuals to believe (place positive probability on) things that are not objectively true. However, it is difficult to square rationality with someone whose beliefs are completely divorced from reality. Therefore, we assume beliefs align with reality at least to some extent.

Condition 1 (Grain of Truth)**.** *For all* $i \in N$, $s_t \in S$, $\mu_i^s(h_t^*) > 0$.

That is, rational individuals do not rule out the true state of affairs. This implies that, although an indivual's beliefs about an event may be wildly inaccurate, that belief is not completely irrational: i.e., for all $W \in \mathcal{H}$ such that $\mu_i^s(W) > 0$, $h_t^* \in W$. Going in the other direction, for all $h_t^* \in H^*$, there exists some $W \in \mathcal{H}$ such that $\mu_i^s(W) > 0$. This condition is not without controversy as it does rule out situations in which an individual is surprised by being confronted with a state of affairs he or she had previously thought impossible. There are formal approaches to dealing with such situations. For now, however, we sidestep such issues.

Learning We can also think of consistencies implied by learning. Even with the Grain of Truth Condition in place, our setup presently allows a person's beliefs through time to be completely inconsistent in all ways except $\mu_i^s(h_t^*) > 0$. For example, suppose $X, Y \in \mathcal{H}$ and $\mu_i^{s_t}(X) = 1$ and $\mu_i^{s_{t+1}}(Y) = 1$ (X and Y contain all the states i believes are possible in periods t and $t+1$, respectively). Then, even if X and Y are quite large, there is nothing in the setup preventing $X \cap Y = h_{t+1}^*$; i.e., the *only* consistency from period to period is belief in the possiblity of the objectively true history. Such situations seem inconsistent with any reasonable concept of learning. The following condition is a notion of learning that admits a wide range of learning models. For example, Baysian updating is consistent with this (though, by no means requred).

Condition 2 (Weak Learning)**.** *Let* $X, Y \in \mathcal{H}$. *For all* $i \in N$, $s_t, s_x \in S, x > t$, *if* $\mu_i^{s_t}(X) = 1$ *and* $\mu_i^{s_x}(Y) = 1$, *then* $Y \subseteq X$.

Notice that learning is, indeed, weak in the sense that one may never learn anything ($Y = X$ through time). However, we imagine that as individuals experience the world, their grasp of it

becomes more refined. Again, this condition is also not without controversy since it seems to rule out "conversion" experiences in which an individual shifts from one worldview to another, apparently inconsistent worldview. Whether or not such experiences are, in fact, inconsistent with Condition 2 we leave for another discussion.

Introspection It seems reasonable to assume that an individual knows his or her own mental features (but may be uncertain of those of others). For example, being certain of one's own beliefs rules out some peculiar mistakes in information processing (e.g., Geanakoplos (1989), Samet (1990)). As described above, the probability distribution representing an individual's beliefs in may vary by state. Introspection entails that, at any given state, the agent's belief assigns probability 1 to the set of states in which he has the same belief as in that state. Formally,

Condition 3 (Introspection)**.** *For each agent $i \in N$ and state $s \in S$, the agent's belief at s, μ_i^s, assigns probability 1 to the set of states in which i has precisely these beliefs: $\mu_i^s(\{s' \in S \mid \mu_i^{s'} = \mu_i^s\}) = 1$.*

Ordering of desires It is also typical to add some structure to desires, namely that they be a partially ordered. Formally, for all $i \in N$, $\preceq_i$ is a partial order relation on the set of paths, P; i.e., the following conditions hold for all paths in Γ:

1. $\forall p' \in S, (p', p') \in D(p)$: the relation ip reflexive,
2. $\forall p', p'' \in p, (p', p'') \in D(p) \wedge (p'', p') \in D(p) \Rightarrow p' = p''$: the relation ip antipymmetric,
3. $\forall p', p'', p''' \in p, (p', p'') \in D(p) \wedge (p'', p''') \in D(p) \Rightarrow (p', p''') \in D(p)$: the relation is transitive.

These conditions simply assume that there is a certain degree of consistency in an individual's desires over states.

Intentions An intention differs from both beliefs and desires in that this mental attitude implies the individual possessing it has made a commitment to take action toward a desired end. The desired end is an event, such as "Mike buys a cup of coffee," which may be actualized by a large number of states of the world; e.g., buying at McDonalds, or at Starbucks, or alone, or with friends, or while believing the dark roast is probably sold out. Thus, in state s, the object of individual i's intention is an event in $\mathcal{S}$. It is not enough for an individual to simply intend some outcome.

Rather, we assume that at the time an intention is formed, it is coupled with a concrete plan of action designed to achieve the desired end.

To formalize this, for each individual i, define an *action plan* as a function $\sigma_i : S \to A$ where $\sigma_i(s) = a_i \in A_i(s)$ indicates that when individual i arrives at state s she selects an act a_i from the set of acts $A_i(s)$ available at that state. Since every state has a single history leading to it, action plans may be history-contingent. Notice that, as defined, the action plan indicates what act the individual will implement at every state. Of course, we do not expect the individual to have thought through a contingency plan for every state in the state space. Rather, we impose a means-ends consistency condition on σ_i that joins the action plan to the intention.

Condition 4 (Weak Means-Ends Consistency)**.** *Suppose individual i's intention is given by* $\gamma_i(s) = X \in \mathcal{S}$*. Let* $P^s_X \subset P$ *denote all the paths in* Γ *that begin at* s *and terminate in* X*. Then* σ_i *is said to be weak means-ends consistent with* $\gamma_i(s)$ *if at no state* s' *along any path in* P^s_X *does* $\sigma_i^{s'}$ *force actualization of a state* s'' *that is not on any path in* P^s_X*. By "force" we mean that* $\sigma_i^{s'}$ *indicates an act that actualizes some state outside of* P^s_X *regardless of the acts of all the other individuals and Nature.*

Condition 5 (Strong Means-Ends Consistency)**.** *Suppose individual i's intention is given by* $\gamma_i(s) = X \in \mathcal{S}$*. Let* $P^s_X \subset P$ *denote all the paths in* Γ *that begin at* s *and terminate in* X*. Then* σ_i *is said to be strong means-ends consistent with* $\gamma_i(s)$ *if at every state* s' *along any path in* P^s_X*,* $\sigma_i^{s'}$ *forces actualization of a state* s'' *that continues along a path in* P^s_X*. By "force" we mean that* $\sigma_i^{s'}$ *indicates an act that actualizes some state on a path in* P^s_X *regardless of the acts of all the other individuals and Nature.*

In other words, Condition 4 says that the individual's plan never has him unilaterally driving the world to a state from which the intended event cannot be reached. When this condition is met, it may nevertheless be the case that the world is driven to such a state. However, this will need to be the result of the acts of others and/or Nature and nothing to do with the acts of individual i. The strong form, Condition 5, says that individual i has a plan of action by which he can gaurantee his intended even regardless of what anyone else does. There is another case which is this: no matter what i does, the intended X will happen. In this case, I do not think we would properly call X intention.

We also need some rationality conditions that tie the preferences over paths to the action plan. This is subtle because paths are determined by the entire act profile (i.e., and not just the acts of

i. So, how do you tie in preferences. One possiblity is to use i's may have beliefs about what the other agents are going to do (remember all of this would be encoded in the states) and, based upon this, choose an action plan that implements the most preferred path possible given the plans of the others. This would then tie beliefs, desires, intentions and plans of action together.

[STOP HERE]

4 Groups

4.0.1 Group composition and existence

Often, we are interested in the individuals that comprise a group. With that in mind, define the *group composition* function $c : M \times S \to \mathcal{N}$ where $c(k, s) = G$ indicates that in state $s \in S$ the group indexed by $k \in M$ is comprised of those individuals whose indices are contained in $G \in \mathcal{N}$. Notice that, using this approach, group composition can differ across states and a given individual can belong to multiple groups in the same state. Indeed, the same collection of individuals can comprise the memberships of different groups; i.e., we can have $c(k, s) = c(k', s)$ for $k \neq k'$.

If k is a potential group in state $s \in S$, then $c(k, s) = \emptyset$. Thus, c maps every element of M (potential or existing) in every state to some element of $\mathcal{N}$ (possibly, $\emptyset$). Yet, because c need neither be injective (one-to-one) nor surjective (onto), the inverse of c need not be implied by c itself. However, we can still define an *inverse group composition* function as $c^{-1} : N \times S \to \mathcal{M}$ where $c^{-1}(i, s) = H$ indicates that in state $s \in S$ the individual corresponding to index $i \in N$ belongs to the groups whose indices are contained in $H \in \mathcal{M}$. We adopt the convention that if s is a state in which i does not belong to any group, $c^{-1}(i, s) = \emptyset$. Then, c^{-1} is a well-defined function that, like c, is neither injective or surjective.

From the preceding setup, we see that a state elaborates all the groups which exist in it. To keep track of this, let $e : S \to \mathcal{M}$ be the group *existence* function $e(s) \equiv \{k \in M | c(k, s) \neq \emptyset\}$. Essentially, e "pulls out of s" the groups that exist in that state. Thus, we can define the "*no-group-exists*" event as $E_\emptyset \equiv \{s \mid e(s) = \emptyset\}$. Assume that S is sufficiently expressive to permit the existence of any combination of groups: for all $H \in \mathcal{M}$, $\exists s \in S$ such that $e(s) = H$. Since states also summarize mental features of individuals, there may be many states corresponding to a particular set of existing groups.

5 Initial conditions

5.0.1 Modest social groups

It appears promising to begin with an analysis of modest social groups and then build to to more complex, formal organizations like firms. Our interest is in *modest social groups*. The conditions required for the existence of a modest social group are stated later. However, we assume that k, contingent upon it existing as a modest social group, has the following informally stated features:

1. It is informally constituted,
2. It consists of two or more individuals,
3. It aims to accomplish a one-dimensional end, and
4. It is one-shot.

This eliminates from initial consideration groups: 1) whose grounding conditions include a concrete explication of group principles (e.g., a contract); 2) which are not singletons; 3) whose purpose is to achieve a single goal (e.g., *take a walk* or *play a duet*, but not *engage in money laundering and kidnapping*); 4) persist beyond the completion or failure of the intended purpose. According to Modest Social Group Condition 2, existing groups have two or more members: $\forall s \in S, c(k,s) \neq \emptyset \Rightarrow |c(k,s)| > 1$.

5.0.2 Analytical sequence

The idea is to begin with the simplest case of an intentional group, one in which the group is constituted simply by its individuals and their relations to each other and the group. Our present interest is in seeing how far we can get in articulating some mutually suitable description of what we mean by group intentions and their associated group acts.

Therefore, assume that the initial state of the world is $s_0^* \in E_\emptyset$, a state in which no groups exist. The profile of mental features is a primitive of the model. Therefore, everyone begins with mental states $\theta(s_0^*)$. These imply a profile of intended actions $a(s_0^*)$. According to these primitives, in a fashion not yet described, some new state of the world, s, obtains in which the groups $e(s)$ come into existence along with the updated mental features $\theta(s)$. Our task is to identify how these all hang together in a coherent metaphysics.

5.0.3 Human acts

To rule out cases of group formation via coercion, like being kidnapped by the mafia and taken to New York in the trunk of a car, we assume that group membership relies upon the classical notion of a *human act*: at the most basic level, $\sigma(s) = a_i$ implies that, in state s, i intends act a_i voluntarily in a fashion "consistent" with his or her desires – i.e., having given his choice some thought and without coercion (we will need to say more about how these features are connected later). One obvious situation that violates this assumption is i finding himself limited to one act at a state s such that $|A_i(s)| = 1$. To avoid this and simplify, assume that, in state s_0^*, all real individuals are free to join any *one* group: for all $k \in M$ and all $i \in N$, $A_i(s_0^*) \equiv \{a_i^{1+}, \ldots, a_i^{m+}\}$.

Note that we have not said anything about the conditions required for group existence. For example individual i intending the act of joining group k, intention $\sigma_i(s_0^*) = a_i = k^+$ is, presumably, necessary but not sufficient to cause a state to arise, s', such that $k \in e(s')$.

5.0.4 Discussion

Although we have still said nothing about how modest social groups come to exist, have group-level intentions or take group actions, we do have the machinery to say a number of things in a precise way. Here are some examples:

1. At s, $i \in N$ knows that the collection of groups Γ exist: $\mu_i(s)(\{s' | H \subseteq e(s')\}) = 1$.

2. At s_0^*, the collection of individuals $G \in \mathcal{N}$ each intend to join group k: for all $i \in G$, $\sigma_i(s_0^*) = k^+$.

3. The event that the collection of individuals $G \in \mathcal{N}$ each intend to join group k: $E_{G \to k} \equiv \{s \mid \forall i \in G, \sigma_i(s) = k^+\}$.

4. In state s_0^*, $i \in N$ knows all the members of G intend to join k: $\mu_i(s_0^*)(E_{G \to k}) = 1$.

5. The *event* that $i \in N$ knows that the individuals G intend to join k: let $\bar{E}_i(s)$ denote the support of $\mu_i(s)$. Then, $K_i(E_{G \to k}) \equiv \{s \mid \bar{E}_i(s) \subseteq E_{G \to k}\}$, where K_i denotes events determined by what i knows in their states. Thus, $K_i(E_{G \to k})$ is the collection of states in which, given μ_i, i knows $E_{G \to k}$.

6. It is *evident* to the individuals G that they each intend to join k: For all $i \in G$, $E_{G \to k} \subseteq K_i(E_{G \to k})$. It can be shown that this implies $E_{G \to k} = K_i(E_{G \to k})$.

7. $E_{G\to k}$ is *common knowledge* at $s \in S$ if and only if there exists an event E such that: $s \in E$ and, for all $i \in N, E \subseteq K_i(E)$ and $E \subseteq K_i(E_{G\to k})$. This is the **?** formulation, which is a restatement of Aumann (1976) in terms of evident events. For example, E can be the event "The individuals G publicly and credibly announce their intention to join k." This announcement is evident to everyone (for all $i \in N, E \subseteq K_i(E)$) and, once it occurs, it implies that everyone knows the individuals G will act to join k, knows that they know, that they know that they know that they know, etc. (for all $i \in N, E \subseteq K_i(E_{G\to k})$). Note that $E_{G\to k}$ is not necessarily evident knowledge: it is possible to have some state $s \in E_{G\to k}$ in which not everyone knows $E_{G\to k}$.

8. In state s_0^*, the individuals G agree that being in k is most desirable: For all $i \in G$ and all $s, s' \in S$ such that $k \in e(s)$ and $k \notin e(s')$, $s' \prec_i s$.

6 Group formation

Since we only have in mind such simple group activities as"we take a walk to NYC" we can think of a fairly simple sequence of acts and consequences that appear to be implied by them. Let us roughly follow (Bratman, 2014, Ch. 2) to see how this setup relates.

Beginning with Section 1, "I intend that we J, and circularity." Let $B \subset N$ be a collection of individuals. For each individual $i \in B$, assume $a_i^* \in A_i(s_0^*)$ is the act that i transports herself to NYC. Let $E_i^* \subset S$ be the event "i is in NYC" and $E^* \equiv \cap_{i\in B} E_i^*$ be the event that all the individuals in B are in NYC. Assume E^* is nonempty and that the members do not start out in NYC: $s_0^* \notin E^*$. Then, the following are some things that Bratman says are *not* a group intention to go to NYC:

1. Each individual in B intends to go to NYC: $\forall i \in B, \sigma(s) = a_i^*$.

2. Each individual thinks being in NYC is the best thing: $\forall i \in B, s' \in E_i^*, s \notin E_i^*, s \prec_i s'$.

Then, Bratman suggests that the key is framing the group intention as "we each intend that we go to NYC." This is where we run into problems because what is being "intended" is vague and, in any event seems to be doing too much lifting. In our framework, an individual can intend his or her own acts – full stop. They cannot intend the intentions or actions of others. In our construction, Bratman's sentence of intention is nonsensical.

While Bratman does indicate that "each of us has the ability to pick out the other participants," [p. 41], I think he leaves out a crucial step: the act of group formation. My sense is that if we make this explicit, we can actually make better headway. The following set of conditions for group formation is incomplete:

1. In s_0^*, the individuals in B jointly intend to bring a group k into existence to go to NYC. This requires several sub-conditions:

   (a) A profile of intentions such that, for all $i \in B$, i intends to join k ($\sigma_i(s_0^*) = a_i^{k+}$) and, for all $j \notin B$, j does not intend to join k: $\sigma_j(s_0^*) \neq k^+$.

   (b) Group existence conditions are now required, such as that the individuals each prefer states in which k contains exactly the individuals B to any other state: for all $s, s' \in S$ such that $c(k, s) = B$ and $c(k, s') \neq B$, $s' \preceq_i s$. The idea is that, since the existence of this kind of group simply requires everyone's assent, i won't remain in the group if the composition is not to her liking. But, to be complete, this needs another condition because we don't know what happens when individuals outside of B also decide to join k. For example, although s is preferred to s', s' may be preferred to any other state. In that case, $c(k, s')$ could, presumably, come to exist.

2. $E_{B \to k}$ (the joint intentions of B to form k) is common knowledge in state s_0^*.

3. Following the intended acts, a new state of the world s occurs in which B forms K: $c(k, s) = B$.

4. In state s, the existence and composition of k is common knowledge.

5. Once the group forms, there must be a plan to get the group to NYC. This is where the idea of group awareness may prove helpful. We may also need to add in structure for planning within groups. This end must be joined to the intentions, beliefs and preferences at play in s_0^* to make everything hang together.

Once the preceding is sorted out, we can start talking about individuals intending and acting from a state of group existence. Thinking about this second part is the next challenge.

6.0.1 Unawareness Structures

References

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics 4*(6), 1236–1239.

Bratman, M. (2014). *Shared agency: A planning theory of acting together*.

Bryan, K., M. D. Ryall, and B. C. Schipper (2021). Value-capture in the face of known and unknown unknowns. *Strategy Science (forthcoming)*.

Dekel, E., B. L. Lipman, and A. Rustichini (1998). Standard state-space models preclude unawareness. *Econometrica 66*(1), 159–173.

Geanakoplos, J. (1989). Game theory without partitions, and applications to speculation and consensus. Technical report.

Harsanyi, J. C. (1967). Games with incomplete information played by "bayesian" players, i–iii: Part i. the basic model. *Management Science 14*(3), 159–182.

Heifetz, A., M. Meier, and B. C. Schipper (2006, sep). Interactive unawareness. *Journal of Economic Theory 130*(1), 78–94.

Heifetz, A., M. Meier, and B. C. Schipper (2008, jan). A canonical model for interactive unawareness. *Games and Economic Behavior 62*(1), 304–324.

Heifetz, A., M. Meier, and B. C. Schipper (2013, jan). Unawareness, beliefs, and speculative trade. *Games and Economic Behavior 77*(1), 100–121.

Mertens, J. F. and S. Zamir (1985). Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory 14*(1), 1–29.

Samet, D. (1990). Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory 52*(1), 190–207.

Schipper, B. C. (2015). *Awareness, in Handbook of Epistemic Logic*, Chapter 3. College Publications.

Schipper, B. C. (2016). Network formation in a society with fragmented knowledge and awareness. Technical report.