Notes on Group Agency[1]

Brian Epstein
Tufts University, Medford

Michael D. Ryall
University of Toronto

February 25, 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

Capital letters (G, N, etc.) refer to sets. Small Arabic and Greek letters refer variously to elements of sets (e.g., $i \in N$), functions (e.g., $\sigma : N \to \mathcal{N}$), and indexed lists (e.g., $x \equiv (x_1, \ldots, x_n)$). Terms are *italicized* at the point of definition. A *profile* is a placeholder for a list of elements; e.g., x where $x \equiv (x_1, \ldots, x_n)$. 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.

3 Individual Agency

Begin with a *population of individuals*, indexed by $N \equiv \{0, \ldots, n\}$ 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 a "Nature" which we assign to population index 0.b

3.1 States

A *state*, denoted s, is a snapshot of the world at a moment in time. States summarize the status *of all features of the world that are relevant to all individuals* 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. The finite *state space at time* t, denoted S_t, with typical element $s_t \in S_t$, contains all the possible states that could occur at t.[1]

A note on terminology: the term 'event', which we define in the next paragraph, is used differently by philosophers than it is by game theorists. We will use the game-theorist's use of the term, but since we are writing to both audiences, it is important to clarify this difference. In game 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. In game theory (and in our usage), an event is a set of states. Because an individual state encodes all relevant features of the world, events provide a way of identifying all the states consistent with a particular feature of interest. For example, in our context, 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.

An *event*, then, is a subset of S. Let $\mathcal{S} \equiv 2^S$ be the set of *events*, with typical element $E \in \mathcal{S}$.

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 to "extract" information from the states. Here, for example, we can let $d(s)$ indicate the die roll outcome in state s. Thus, given $s \in S$, $d(s) \in \{1, 2, 3, 4, 5, 6\}$, assuming the die never lands on an edge. To clarify the domain and the range to which the function maps, we write $d : S \to \{1, 2, 3, 4, 5, 6\}$. Now, the event "the die roll is even" is described by $E \in \mathcal{S}$ such that $E \equiv \{s \in S | d(s) = 2, 4 \text{ or } 6\}$.

3.2 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 changes that, in addition to the acts of the individuals, determine actualization of a particular state from an immediately preceding, previously actualized state). We assume the world begins in a null state, denoted s_0. For all individuals $i \in N$, $A_i(s)$ indicates the set of *feasible acts available to* i *in state* s with typical

[1] This setup can be generalized to include countably or uncountably infinite state spaces. Limiting attention to finite sets allows us to sidestep some mathematical complexities which are not necessary for the purposes of this analysis.

element $a_i \in A_i(s)$.[2]

We adopt the convention that $A_i(s) = \emptyset$ indicates that individual i has no available acts in state s. An *act profile* is a list of acts, one for each individual, denoted $a \equiv (a_0, a_1, \ldots, a_k)$. Recall, Nature is "Individual 0" so that a_0 summarizes all the relevant natural developments that, in conjunction with the individuals' acts, determine which state is actualized following s. The set of *all act profiles at state s* is $A(s) \equiv \times_{i=0}^{n} A_i(s)$. The set of *all possible act profiles* is $A \equiv \cup_{s \in S} A(s)$.

3.3 Dynamics

As indicated above, the act profiles summarize all the contingencies required to actualize one state from the next. To formalize this, let $\omega : S \times A \to S$ be the *state-contingent actualization function*, where $\omega(a_t|s_t) = s_{t+1}$ indicates that if the act profile at state $s_t \in S_t$ is $a_t \in A(s_t)$, then the next state actualized is s_{t+1}. Assume that, for all s_t, $\omega : \{s_t\} \times A_t(s_t) \to \cup_{a_t \in A_t(s_t)} \omega(a_t|s_t)$ is a bijection. In other words, each feasible act profile in a given state leads to a unique state and each state can be traced back to its immediate predecessor by the unique act profile that would lead to its actualization.

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. One way of thinking about this is to imagine that a state also encodes the distinct history of acts that lead to it.

The world begins at state s_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. That is, $a_0 = (a_0, \emptyset, \ldots, \emptyset)$, where a_0 represents all the factors that lead individuals to their first decision state, $s_1 = \omega(a_0|s_0)$. 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" prior to the start of the focal time period.

We define the *history at state s_t* as a profile that starts at s_0 and ends at s_t along with the acts that determine the sequence of states according to ω, denoted $h(s_t) = (s_0, a_0, \ldots, a_{t-1}, s_t)$, where $s_1 = \omega(a_0|s_0)$, etc. The set of all histories at time t is H_t. Let T be the number of time periods

[2] Notice that we use a capital letter to indicate that $A_i(s)$ is a set-valued function. Also, 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.

under consideration for the analysis. Then, the set of all possible histories is H_T and the set of all subsets of histories is $\mathcal{H}_T$. An arbitrary *history at time t* is denoted $h_t \in H_t$, where we start with the *null history* $h_0 = (s_0)$ at the beginning of time (so, $H_0 = \{h_0\}$ and $S_0 = \{s_0\}$). Because there is a single root node and ω is a bijection, the set of paths in H_T form a tree. Thus, S can be partitioned according to subsets of states corresponding to time periods: $S = S_0 \cup S_1 \cup \cdots \cap S_T$ and $S = S_0 \cap S_1 \cap \cdots \cap S_T = \emptyset$. Note also that each S_t implies a partition of H_T according to the sets of paths intersecting the states in S_t.

3.4 Mental attitudes

In what follows, we develop a belief-desire-intention model. As we elaborate below, we assume that the status of one's mental attitudes at time t is a feature of the actualized state s_t. First, we define *beliefs* as subjective conjectures about the likelihood of past events, the present state, and future events. Second, *desires* are the individual's attitudes toward histories, represented as a partial order relation on $\mathcal{H}_T$. In other words, individuals consider both the sequences of states they experience as well as the acts that induce them. For example, even though a junior faculty member may have the same degree of desire for tenure at her present institution in all states of the world in which that happens, some paths to tenure may be more costly than others depending upon the acts required to get there (both her own and others, including those by Nature). Making the desire relation a partial order on $\mathcal{H}_T$ implies a partial order on H_T but also allows individuals to compare events in $\mathcal{H}_T$ (e.g., the event of gaining tenure vs. not). Third, *intentions* represent an agent's commitment to undertake a plan of action designed to actualize an event in $\mathcal{H}_T$. As we will see, others' perceptions of one's intentions will play a social role in our framework.

++

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.

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.

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 $\theta_i \equiv (\mu_i, D_i, \gamma_i)$.[3] 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.

3.5 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 $h_t^* \in H^*$,

[3]In setting up mental features in this way, we are following a version of the familiar "type-space" approach used in game theory (See **??**).

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., **?**, **?**). 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 posessing 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 $P_X^s \subset P$ 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_X^s does $\sigma_i^{s'}$ force actualization of a state s'' that is not on any path in P_X^s. By "force" we mean that $\sigma_i^{s'}$ indicates an* *act that actualizes some state outside of P_X^s 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 $P_X^s \subset P$ 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_X^s, $\sigma_i^{s'}$ forces actualization of a state s'' that continues along a path in P_X^s. By "force" we mean that $\sigma_i^{s'}$ indicates an act that actualizes some state on a path in P_X^s 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

4.1 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

5.1 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: $\forall s \in S, c(k,s) \neq \emptyset \Rightarrow |c(k,s)| > 1$.

5.2 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.

5.3 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 $k \in e(s')$.

5.4 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 **?** 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 (**?**, 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.

6.1 Unawareness Structures