Examples of Awareness of Intentions[1]

Brian Epstein

Tufts University, Medford

Michael D. Ryall

University of Toronto

July 10, 2021

---

[1]Rough examples for the paper

We consider the problem of Brian's Infant, whom we refer to as individual i. At time $t = 0$, the i finds himself in a room with two toys, A and B, situated in front of him. At this point, i needs to figure out what to do. His goal is to obtain the best toy.

Parsimonious game theoretic treatment

The parsimonious game theoretic treatment of this situation is shown in Fig. 1. The uncertainty of indiviudal 1 with respect to which toy is actually best is represented by including an initial move by Nature at time $t = 1$ – i.e., the two possible states are A-Best and B-Best, one of which is true and the other of which is counterfactual. The bold lines indicate the choices of the players. Individual 1 is then faced with a choice as to whether to get A or get B. As illustrated by the dashed line connecting 1's decision nodes, 1 does not know with certainty which is the true state but, as indicated, believes it is most likely that A is best. Therefore, 1 chooses to get A in $t = 2$. As a result, 1 obtains A in $t = 3$.

The features to note are: 1) the world simply presents 1 with a decision; 2) although 1 is uncertain about which toy is best, he is aware of the counterfactual possibilities – indeed, 1 knows everything about the game, including what will happen as a consequence of his actions; 3) all of 1's cognitive processes associated with the decision are compressed into the act of making a decision. The decision could be elaborated as one involving probabilistic beliefs on the part of 1, but this is not necessary. For whatever reason, at the time of his decision, 1 believes (with some measure of uncertainty) that A is best. Given these beliefs, and a desire for possessing the best toy, 1 chooses to get A.

State spaces, fields of awareness, and awareness structures

A *state* captures the world at a moment of time. The state space in a past or present period contains the state that actualized at that time and all the counterfactuals (states that could have occurred instead at that time). The state space for a future period contains all the states that could occur at that time.

In this example, the simplest state space is $S_0^n = \{A^*, B^*\}$. The n superscript indicates that the state space is associated with "Nature" (i.e., these states are the ones associated with the world as it really is). Here, either Toy A or Toy B is the best toy for i. Which is truly the best is determined
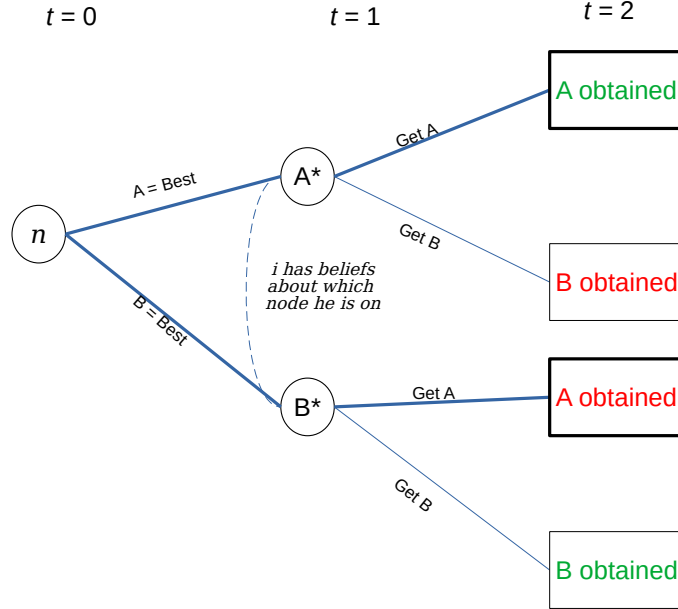
Figure 1: Brian's Toddler, parsimonious game-theoretic treatment

by Nature at the beginning of time.

The toddler lands in the situation at $t = 0$ in one state or the other. States elaborate everything about the world at a given time, including the cognitive status of everyone involved. Here, we wish to expand the parsimonious representation in Fig. 1 to dive more deeply into i's cognitive status as it evolves through the decision process.

We are interested in describing what i knows (or would have known) about the actualized (or counterfactual) states of the world contained in S_0^n. We do this by equipping i with a description of the world as he is aware of it. We refer to this description as i's *field of awareness* (FOA). Because i's cognitive status is determined by the actualized state, we write, e.g., $S_0^i(A^*)$ to indicate the state space describing i awareness in state A^*. If i is fully aware, then $S_0^i(A^*) = S_0^n$. Specifically, the FOA includes everything about the actualized state about which i can reason, including counterfactual states.

In addition, the state elaborates i's beliefs about which state is true. For example, it may be true that A is the best toy (state A^*) and that i knows it. Alternatively, it may be true that A is best, but i is uncertain about it. Fig. 2 panel (a), illustrates the case in which A is the best toy
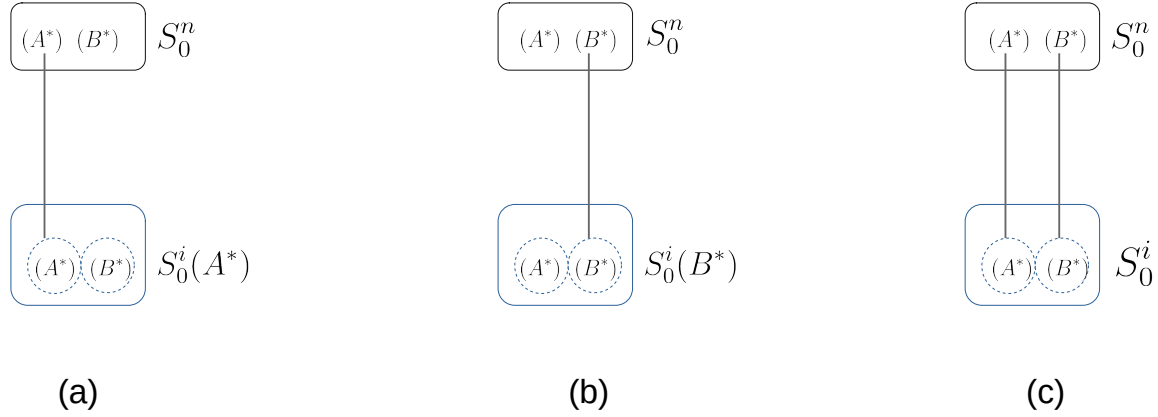
2

(a)   (b)   (c)

Figure 2: Toddler's FOA in (a) state A^*, (b) state A^*, and (c) shorthand, combined diagram

(shown by brackets around A^* in space S_0^n). Below S_0^n we depict $S_0^i(A^*)$. The grey line shows that, when A^* is true, the FOA for i is $S_0^i(A^*)$. The toddler is aware of both states. The dashed lines around the states indicate what he knows. Here, he knows that the true state is A^*. Moreover, he can reason about counterfactual state B^*. The blue dashed line around B^* means that i believes that, had B^* been the case, he would know it.
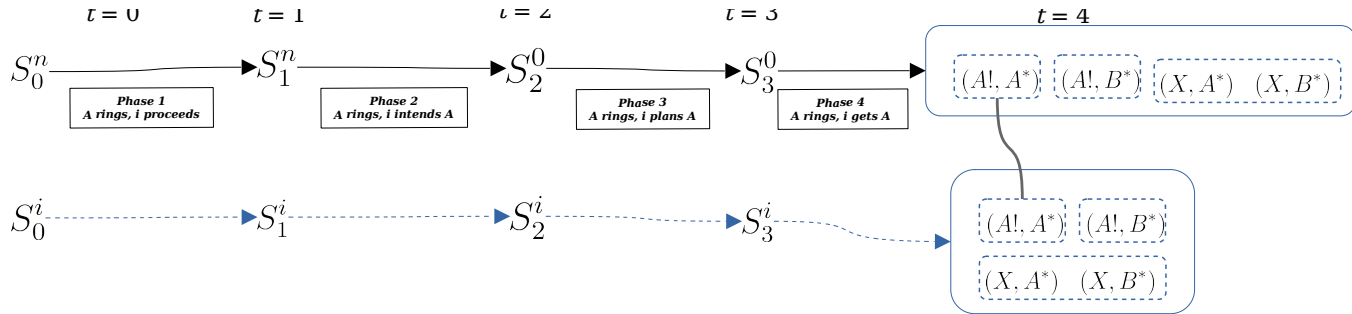

Figure 3: Toddler's FOA in (a) state A^* with uncertainty about which state is true

Panel (b) illustrates the FOA that arises when B^* is the true state of the world. When awareness and beliefs are consistent across states, diagrams can be combined without danger of ambiguity, as shown in Panel (c): when A^* is true, i knows it and, moreover believes that, had B^* been the case, he would have known that ... and the converse when B^* is true. Illustrated this way, the

3

implication is that beliefs about which toy is best are the same in both state. In other words, Nature determines which toy is best, i sees the two toys and forms beliefs about which is best. These are *the* beliefs regardless of which toy is truly best. Note that this diagram corresponds to the information set (the nodes connected by the dashed line) illustrated in the parsimonious game shown in Fig. 1.

Fig. 9 Panel (b), illustrates a situation in which i is unaware that Toy B could be the best. In this case, i's FOA includes a single state: A^*. The diagram shows both states projecting into it. That is, whichever state is true, i is only thinking about Toy A. Perhaps Toy B is under the couch or, even if it is in i's field of vision, he simply isn't thinking about it. Not that, in the situation depicted in Fig. 9 Panel (a), i could believe that B^* is impossible – i.e., a probability-zero occurrence. In Fig. 9 Panel (b), the possibility of B^* is not even in his mind – he is completely unaware of it. As we show elsewhere, these two cases (zero-probability and unawareness) are *not* equivalent.

Finally, Fig. 9 Panel (c) illustrates the flexibility of the setup. Here, in state A^* the toddler is uncertain whether A^* or B^*. However, in state B^*, he is certain that B^* is true and reasons (incorrectly) that he would similarly be certain that A^* had that state occurred. Given the simplicity of the story, it is hard to provide an explanation of why this sort of inconsistency would arise. Nevertheless, the framework is sufficiently general to accommodate it.

A four-phase decision process

Our goal is to expand the parsimonious treatment of individual i's cognitive process to include some of the elements discussed in the philosophy literature. We begin by elaborating what i is aware of about the state of the world, how he thinks about his decision at a given moment in time, how the thinks about his decision in light of possible future states and how this evolves dynamically over time.

First, what is the decision process? One useful disaggregation for our context is one that involves four-phases. At $t = 0$, individual i finds himself in a state of the world in which Nature presents him with the potential to make a decision about something. For example, our toddler finds himself faced with a potential welter of information along with an opportunity to begin the process of achieving some goal. For example, he is set down on the with the two toys in front of him. Toy A is making a ringing sound. There are other people in the room doing interesting things, the TV is

on, and the pet dog is barking in the next room. The thought occurs to him that one end he might consider is obtaining one of the toys. The toddler chooses to think about obtaining the best toy. In this setting, he naturally organizes the available information for the next phase – he considers the condition of the toys, realizing that what the other people are doing, what is happening on the TV, and what the dog is doing are all irrelevant to an analysis of which toy is best.

Importantly, *states* indicate what is, could have been, or could be true about the world at a moment in time. Alternatively, *phases* represent the flow of the world between states. In this framework, phases are the periods within which acts occur. Acts include the acts of Nature, physical actions by individuals, and non-physical acts by individuals (e.g., pondering a choice). Acts are the events that cause the world to evolve from one state to the next over time (where "Nature's acts" are all the things that co-determine the next state in conjunction with the acts of individuals). It is also important to note that the acts associated with higher phases require the completion of the lower phases. For example, one cannot begin analysis until one chooses to do so and organizes the information required for it. In game-theoretic language, the acts associated with Phase k are not *feasible* until Phases 1 through $k-1$ are completed.

At the conclusion of Phase 1, the decision maker arrives in $t = 1$. The actualized state in which he finds himself is determined by the state actualized in $t = 0$, the acts of Nature, of i, and of others during Phase 1. At this point, i has either settled upon opting to continue observing the situation further (which begins another iteration of Phase 1) or to move forward and analyze whether to form the intention of interest based upon the information organized in Phase 1 (which begins Phase 2).

In **Phase 2**, i analyzes the information organized in Phase 1 and determines whether forming an intention is, indeed, warranted and, if so, which one specifically. Intentions are commitments to achieve some future state of the world (i.e., in which the desired end is obtained). For example, in Phase 1, the toddler chose to think further about forming an intention to obtain the best toy. As part of that choice, he also organized the available information so as to support that analysis (the condition of the toys versus other features of his environment). Suppose that, at the beginning of Phase 2 ($t = 1$) Toy A is still making the ringing sound and that this leads the toddler to conclude that A is best. As Phase 2 ends (at $t = 2$), the analysis is concluded and an intention is either formed (move to on Phase 3), the analysis is chosen to be extended (reiterate Phase 2), or the movement toward the focal end is abandoned (return to Phase 1).

Phase 3 begins at $t = 2$. During this phase, i formulates a plan to complete the intention. A plan provides a decision maker with a menu of state-contingent acts over time designed to achieve the intention. In a given state, the decision maker is presented with information about the present state, beliefs about both the present and the future (and, thus, beliefs about what the acts of Nature and others will be), and a set of feasible acts to do in during the next phase. A rational decision maker chooses a plan that selects the best among the feasible acts at each future state (with respect to achieving the intended end).

Here, it may be helpful to make some comparisons with game theory. In game theory, what we are calling a *plan* is referred to as the decision maker's *strategy*. Typically, game theory presents an interactive decision situation (i.e., in which players' decisions interact to affect the outcome of the game) as an extensive for game. Fig. 4 expands Fig. 1 to depict a situation in which the toys ring and the toddler is unsure of which toy is best. Here, the information sets (nodes connected by dashed lines) indicate that, at the time of his decision, i knows which toy is ringing but not which is best. However, i's beliefs are that the ringing toy is best. Then, in this game, Nature determines the state, i forms beliefs based upon what is happening, and chooses rationally.

Here, i's rational strategy is: If A is ringing, get A, otherwise get B. Note well that "rationality" does not imply perfect foresight. Rather, it means that i chooses the subjectively optimal action given his beliefs. Beliefs may or may not be consistent with objective reality, depending upon the primitives of the game. In our example, a well-calibrated Bayesian player would believe that Nature selects among the four possible states according to some probability distribution. His beliefs about which node is really best would equal the true conditional probability of which is best given which toy is ringing. In most games, a players' desires are represented by a dollar-denominated utility number associated with each path through the game (the terminal nodes indicate the outcome of play). Then, the rational Bayesian player selects the action that maximizes his expected utility. More generally, players' subjective beliefs need not be consistent with objective reality, in which case the "rational" player selects according to his subjective expected utility.

This brings us to the notions of *complete* versus *perfect* information games. A game of complete information is one in which the players know everything about the structure of the game, including everyone's payoffs: that is, the players know what game they are really playing. Games of perfect information are those in which, at every decision node, the associated player knows everything that happened in the past (i.e., there is no uncertainty regarding which node each player is on at any
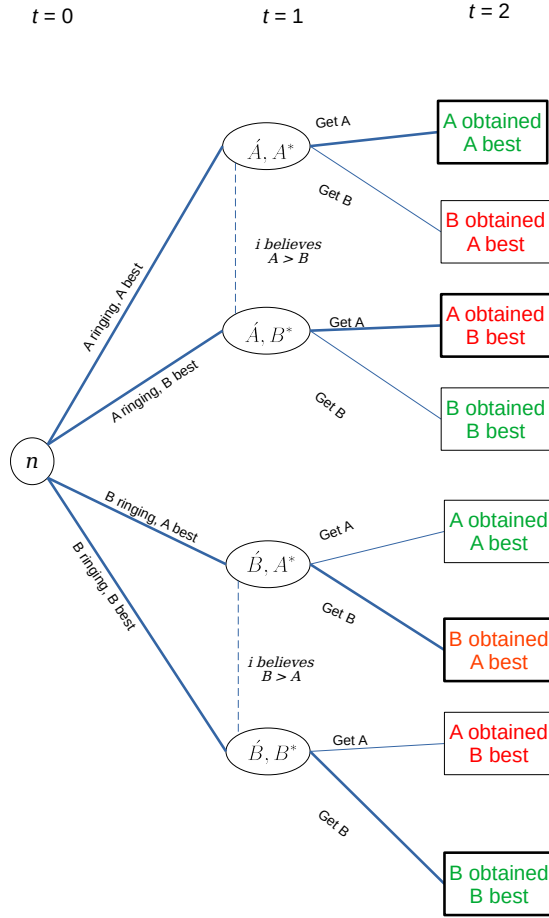
6

Figure 4: The complete but imperfect information game in which Brian's Toddler must act

point). In Fig. 4, the assumption is that Brian's Toddler knows everything about the game, but at the time of his choice of action, is uncertain about which toy is best. Games of incomplete information are typically analyzed by transforming them into games of imperfect information: players know the set of possible games, Nature chooses the true game at the beginning of time and, faced with a decision, the player's form beliefs about both which game they are playing and where in that game they find themselves. The game in Fig. 4 is a game of complete but imperfect information.

The parsimonious game theoretic treatment focuses upon the actions of the decision makers. In Fig. 4, Nature acts by choosing the true state, i arrives at a decision about which toy to choose

given uncertainty about which is best. His choices are A or B. Following his choice, the game resolves with an outcome. The question being asked in Fig. 4 is: *suppressing the details* of how he comes to formulate his strategy in the depicted situation (but making some assumptions about rationality), what strategy would that be? The treatment is "parsimonious" because the Phases 1-3 in our framework are left implicit.

In **Phase 4**, i arrives at $t = 3$ and – if the state at that point is consistent with the plan developed in Phase 3 – acts in accordance with it. One way a state would be inconsistent with the plan is if it was not contemplated as a possibility in the plan's "menu" of possible states. (Such inconsistencies are not possible in a complete information game.) We will have more to say about what constitutes an inconsistency later but, for now, *we claim that running into an inconsistency causes the decision maker to revert to Phase 1.*

Detailed illustration

We begin by illustrating the case in which everything goes smoothly. **Phase 1:** i hears A ringing and decides to move forward to Phase 2. **Phase 2:** Toy A continues to ring and, influenced by this information, i believes that A is best, thereby forming an intention to get A. **Phase 3** A continues ringing and i plans to effectuate his intention by crawling to the location of Toy A and picking it up. **Phase 4** A continues ringing and i crawls to Toy A and picks it up, thereby concluding the decision-action process. Notice that i's intention to obtain the best toy may or may not have been satisfied at $t = 4$. If so, the process concludes. If not, i may choose to think about obtaining B.

The situation in $t = 0$ is illustrated in Fig. 5. Phase 1 generally involves limiting the Phase 2 FOA to a subset of focal states. Thus, given all the elements of the world of which i is passively aware at $t = 0$ (other people, the TV, and the barking dog), he will limit his attention to a strict subset of states thought to be decision-relevant. In our depiction, the decision-irrelevant features of reality mentioned above are not present at $t = 0$ (we erase them from S_0^n to simplify the diagram – but the reader is welcome to imagine a much larger state space at $t = 0$ that includes all these elements).

Specifically, we define the states of the world at $t = 0$ as $S_0^n \equiv \{(\acute{A}, A^*), (\acute{A}, B^*), (\acute{B}, A^*), (\acute{B}, B^*)\}$ where the accent indicates that the toy is ringing a bell: e.g., $(\acute{A}, A^*)$ is the state in which A is ringing and A is best. Assume this is the true state. Further, assume i is fully aware but uncertain which toy is best. This is indicated by the blue box, and labelled S_0^i. As shown, i is aware of all
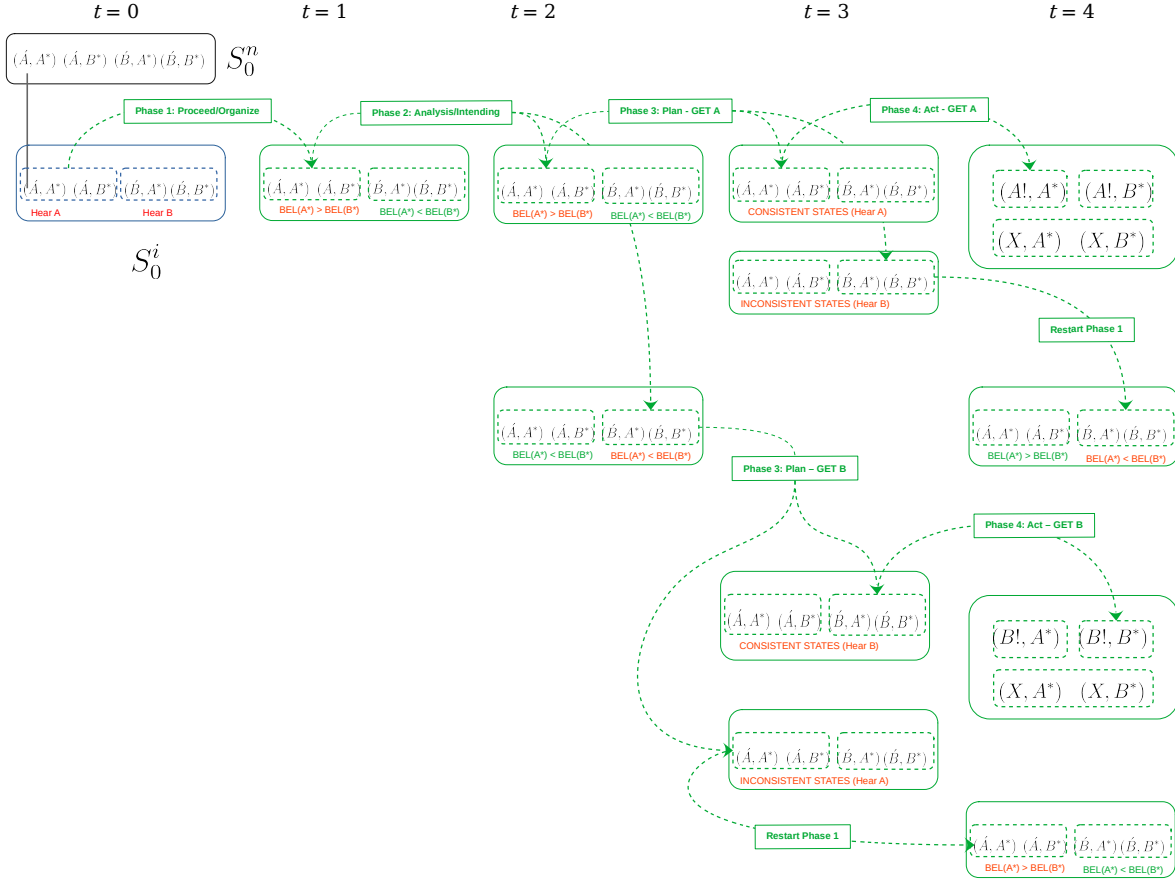
Figure 5: The four-phase process as considered by i at $t = 0$

the states and knows A is ringing. However, as shown by the dashed lines (*information sets*), he is uncertain whether A^* or B^*. The solid grey line shows that the FOA depicted is the one that arises in state $(\acute{A}, A^*)$. At $t = 0$, i's has some belief endowment.

In addition to what the individual is aware of with respect to the present situation, he also has thoughts about the future. Tracking all four phases, our individual considers the future unfolding as shown by the green objects. Moving to Phase 2, the FOA may be reduced or expanded, as discussed, and Nature may act in a way that changes beliefs. Once the analysis is complete, and the intention is formed to obtain the preferred toy, at which point i formulates a plan.

To emphasize, while the individual is thinking, analyzing, planning, and acting, the world evolves. Our individual i considers how things might unfold. For example, during the analysis, which ends at $t = 2$, new information may come in to change i's assessment of which toy is best.

9

During this interval, a plan to get A or to get B is being shaped. The planning phase follows the intention. In this example, i is rational – he plans to get whichever toy he believes best. In this case, real-world events may once again intrude upon the process, this time in ways that disrupt the plan. This is illustrated in the figure. For example, i may plan to get A yet experience something that happens to indicate that B is really best (e.g., Toy B starts ringing). When the plan is disrupted, we assume that the decision maker must backtrack to the analysis phase.

Then, if the state of the world in $t = 3$ is consistent with the plan, i proceeds to act. So, in the top row of individual 1's projected future, following the plan to Get A, a state of the world occurs in which 1 continues to believe that A is best and, hence, 1 acts to obtain A. However, i considers the possibility that he will land in an inconsistent state in $t = 3$.

If the FOA evolves in a fashion consistent with the plan, then i moves to the final phase in which he acts to obtain A. When a toy is obtained, the state is indicated so with an exclamation mark (e.g., $A!$). Here, i allows for the possibility that he fails to obtain his planned objective, indicated by X instead of $A!$ or $B!$. Note also, that i believes that, upon successful completion of the action to get a toy, he will know whether it is the best one or not (indicated by the green dashed lines) and will be able to reason about the counterfactual (obtaining the planned toy but realizing it is not the best). If something happens to disrupt the plan, i will be uncertain about which is best.

Suppose everything goes according to expectations. Fig 6 shows the situation at $t = 1$. The world has evolved to S_1^n in which state $(\acute{A}, A^*)$ continues to hold. Reality is illustrated on the top row. Individual i begins the analysis phase based upon the information organized in Phase 1.

Keep in mind that the state labelled $(\acute{A}, A^*)$ in S_1^n is not identical to the one so labelled in S_0^n, even though the state spaces appear to be the same. First, the states in S_1^n include the information about the sequence of preceding states (i.e., $s_1^n = (\acute{A}, A^*)$) as well as about the acts that caused them (i.e., A *continues* ringing and i chooses to consider forming an intention in Phase 1). Second, individual 1's state of mind has changed (and, remember, this is also summarized by the state). He recalls what he knew before, S_0^i as well as what actions were taken by himself and Nature. This is indicated by the blue dashed line. Another difference is that he projects the decision process from the present into the future. This is indicated by the green objects – although they are consistent with his state of mind in $t = 0$, they are what is in his mind in $t = 1$.

In Fig. 7, Phase 2 concludes in $t = 2$ with i forming an intention to get A. Toy A continues to ring. This is relevant because we have assumed that i believes whichever toy is ringing is best.
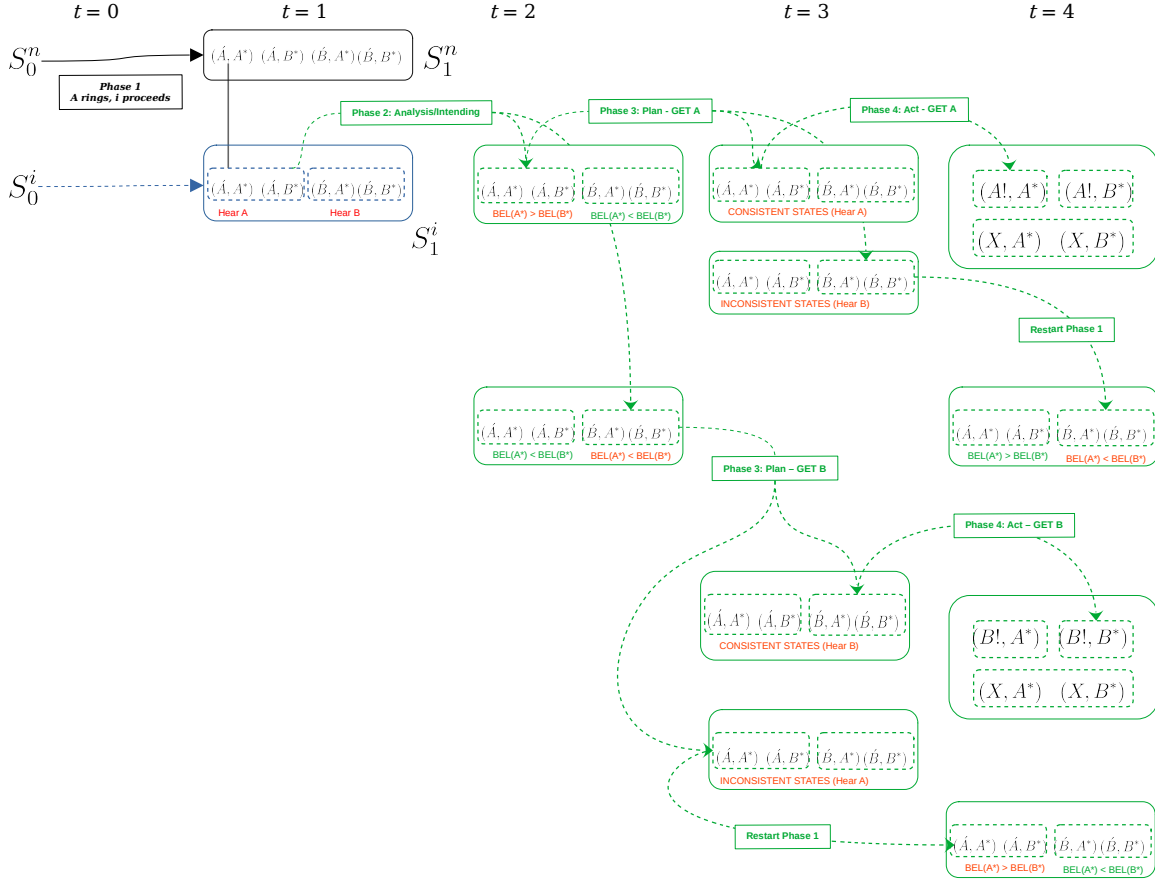
10

Figure 6: Environment stable, Brian's Toddler begins analysis,

Since A rings throughout, there is nothing happening on the part of Nature to disrupt the process.

In Fig. 8, Phase 3 concludes in $t = 3$ with a plan formulated to act towards the attainment of the intention. The sound of A ringing continues. This is consistent with the plan to get A. Therefore, i proceeds to act according to plan – that is, to actually get A.

Finally, as shown in Fig. 9, the process draws to a conclusion: toy A is obtained and, as it turns out, this is indeed the best. Notice that all uncertainty has been resolved – by obtaining A, it is discovered to be the best.
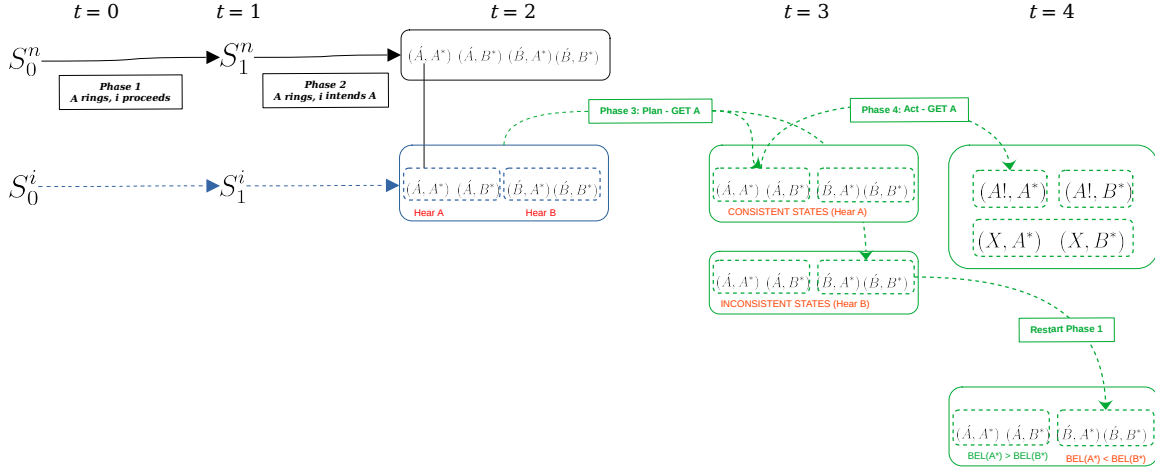
11

$t = 0$ $t = 1$ $t = 2$ $t = 3$ $t = 4$

S_0^n — Phase 1 / A rings, i proceeds — S_1^n — Phase 2 / A rings, i intends A — $(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$

S_0^i — S_1^i — $(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$ — Hear A / Hear B

Phase 3: Plan - GET A   Phase 4: Act - GET A

$(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$ — CONSISTENT STATES (Hear A)

$(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$ — INCONSISTENT STATES (Hear B)

$(A!, A^*)\ (A!, B^*)\ (X, A^*)\ (X, B^*)$

Restart Phase 1

$(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$ — BEL(A*) > BEL(B*) / BEL(A*) < BEL(B*)

Figure 7: Environment stable, intention formed, planning begun

$t = 0$ $t = 1$ $t = 2$ $t = 3$ $t = 4$

S_0^n — Phase 1 / A rings, i proceeds — S_1^n — Phase 2 / A rings, i intends A — S_2^0 — Phase 3 / A rings, i plans A — $(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$

S_0^i — S_1^i — S_2^i — $(\acute{A}, A^*)\ (\acute{A}, B^*)\ (\acute{B}, A^*)\ (\acute{B}, B^*)$ — Hear A / Hear B

Phase 4: Act - GET A

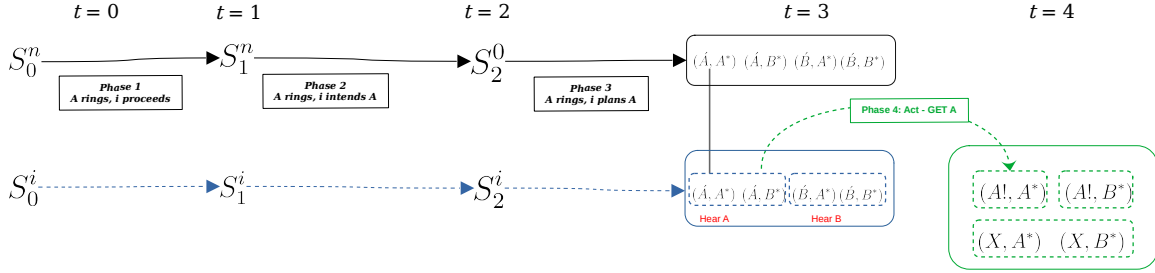$(A!, A^*)\ (A!, B^*)\ (X, A^*)\ (X, B^*)$

Figure 8: Environment stable, plan complete, action begun

Can this be represented as an extensive-form game

So far, we have not written anything down that can't be shown as an extensive form game. This is not surprising given that, thus far in our example, individual i continues to have full awareness of the situation and, as such, has expectations about the future that are consistent with reality. To represent what we are doing as an extensive form game, Fig. 4 would need to be expanded to include all the acts of i (organize information, form intention, set up plan, act to get a toy, etc.) as individual decisions. As, well, it would need to account for the acts of Nature in each intervening period (choosing a ringing toy).

Unlike an extensive form game, our diagrams provide an elaboration of what is going on in
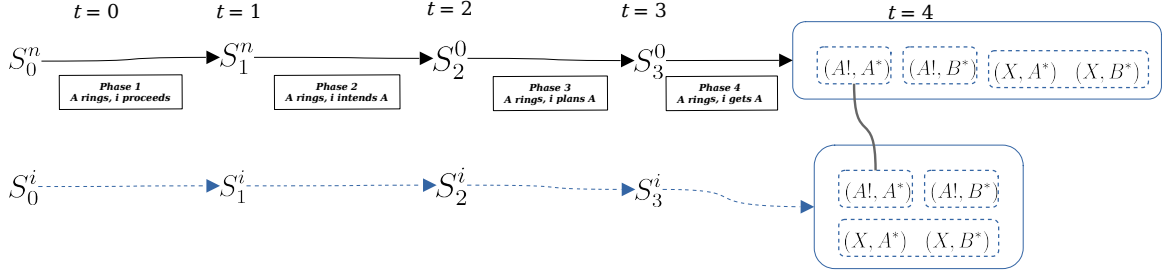
Figure 9: Process complete, intention satisfied!

i's mind (explicitly: recalled history, FOA, and projected futures). However, this level of detail comes at the expense of illustrating everything happening over time in one diagram (.e.g, Fig. 4). Game theorists would surely point out that, while these diagrams are not incorrect, they are also not necessary. Elaborating all these details in our example is superfluous since they are all implied by the joint stipulation that the setting is one of complete information and perfect recall (i.e., i's assessments about present and future possibilities is correct and he remembers everything he knew at in an earlier period). Nevertheless, since this paper is attempting to bridge idea between two disciplines, we judge the additional elaboration to be worthwhile.

The next step is to illustrate what can go wrong – how an even in a complete information setting, a decision maker can get stuck in a "paralysis by analysis" situation. This, then will bring us to introduce our idea of intentional unawareness and its effects on the decision process.

Paralysis by analysis

The previous discussion suggested that things can go wrong when nature evolves state spaces more quickly than the decision maker's response process. This situation is illustrated in Fig. 8: the interpretation now is that the state actualized in period $t = 2$ reflects an act by Nature (not illustrated but, e.g., Toy B sounding a bell) that interrupts the intended plan by causing the toddler to switch beliefs from $A > B$ to $B > A$. Given this new state of mind, a new plan is formulated – to get B.

This alternating flow of information can go on indefinitely, as illustrated in Fig. 10. Under the standard, Bayesian belief-desire model, a rational agent is one who immediately updates beliefs
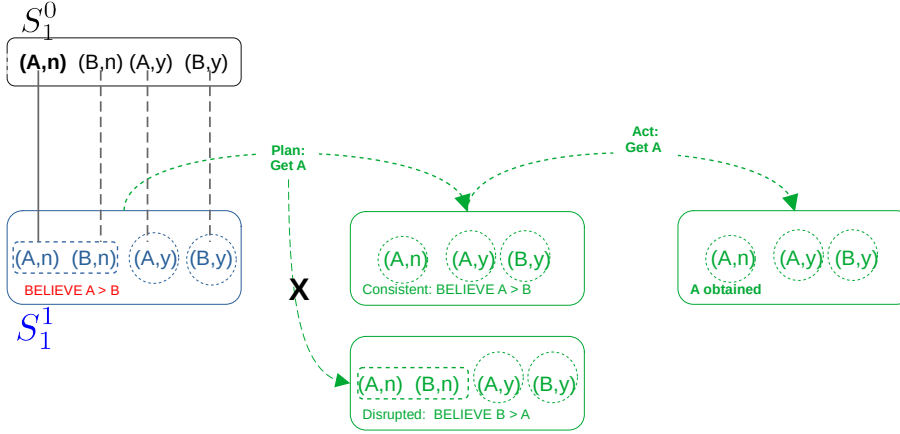
Figure 10: Nature disrupts the plan

based upon new information. This is true in the sense that, provided the information is indeed informative (and not just noise), more information will lead to better decisions. The problem highlighted here is that taking new information into consideration requires an allocation of time and cognitive resources, both of which are in finite supply.

By explicitly accounting for this fact, we see that a truly rational decision maker must weigh the benefits of recalibrating while postponing acting versus ignoring new information and moving forward. The purpose of making a decision is to act and of acting is to achieve an end. If one never decides, then one attains the desired end – the means to which is the deciding. (Note that "never" is too high a bar – if one discounts future utility streams, then there is always a tension between taking time to analyze in lieu of acting.)

Intentional unawareness

We can now see how unawareness solves the problem. Here, there are three interrelated aspects: intention, planning, and awareness. How these are distinguished from one another in a fashion amenable to philosophers is a question that requires further discussion. Speaking roughly, at some point, an individual "commits" to making an act.

It seems that an intention (a commitment to attain some end) always implies a corresponding
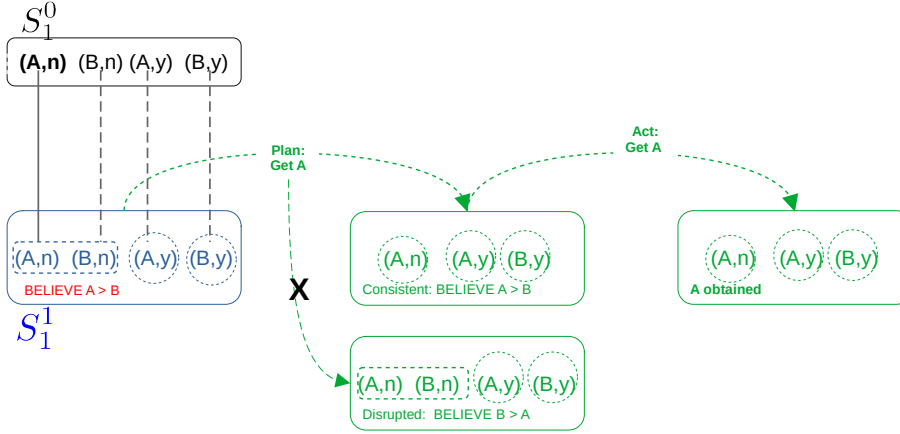
Figure 11: Nature disrupts the plan again

commitment to a plan (a program of action designed to achieve the intended end). Even in the case of split-second decision making, e.g., a policeman's intention to stay alive during a drug raid upon observing an unidentified person moving in the room behind the door, a "plan" is required – e.g., aiming the rifle and pulling the trigger. Thus, a random twitch in one's leg does not count as an intentional act. The converse is not true: one can make many plans without intending to put them into action.

The new idea that we are proposing is that along with the plan comes another piece of the puzzle – intentional unawareness. In the most disaggregated elaboration, there are at least two places where unawareness arises. The first is at the observation/analysis stage. Here, the individual decides what aspects of the present state of the world to pay attention to (we will call this *weak* unawareness). Having focused upon a particular set of aspects about the world, the individual proceeds to conduct an analysis. The conclusion of the analysis is a decision either to continue thinking about and analyzing the situation or to commit to the attainment of some end.

In the preceding paragraph, I mention weak unawareness because it seems there is an important distinction worth making. One can be unaware of some things which one could call to mind (weak) as well as of some things about which one cannot call to mind (strong). As I was writing a moment ago, I had music playing in the background. I was unaware of the name of the band playing the

music. It was not in my mind at all. I was not thinking about it. Then, as I started thinking about examples to write down, this one occurred to me. When it did, I naturally recalled the name of the band. A moment ago I was also unaware of what engineering details are required for a working teleportation system (which could be the null set if such systems are impossible). Now, I am aware of the question, but not of the answer – nor will I ever be.

Here, there is another distinction. A moment ago, I was not thinking about what the temperature is in Hong Kong. It was not in my mind at all. Now, the question is in my mind – I am aware of it. In this case, I have introduced a new information set into my FOA. It includes a *range* of temperatures within which I think the true temperature lies. Presumably, I also form beliefs over that range and can report an expected temperature. I can also take an action (look on the internet) to discover the actual state of the world.

All of these awareness distinctions seem important to intending and planning.

With this commitment in place, the individual then develops a committed plan of action. The program may be simple, i.e., just selecting among one's presently available acts. Or, it may be complex, requiring much analysis – including deciding the best among several plans required to attain the end. In any event, the conclusion of this process is the committed plan. Unawareness arises here because a plan can be enacted without further analysis. An important caveat is that there must be some specification of what states the plan are included in the plan's FOA with the proviso that should a state arise that is not part of the plan's FOA – that is, should the individual become aware of something that falls outside the plan FOA, then the plan is disrupted. In other words, the plan allows the individual to put activity on "auto-pilot" for some FOA. Awareness of new states or events outside the scope of the plan have the potential to disrupt it – at a minimum, a reevaluation is required. [This bit needs further thought and refinement.]

With all of this in mind, return to the problematic case discussed above. Fig. 12 illustrates the situation in $t = 1$ in which the intention, plan, and planned unawareness all occur in the period. Notice the change from Fig.XXX: now, instead of a more refined set of future FOAs, the (B, n) state has been eliminated from the future plan. The shift is intentional (or, at least, is implied by the plan). The potential for state (B, n) to disrupt the plan is eliminated.

The implementation of the plan is shown in Fig. 13. The world evolves and, once again, Nature does her best to disrupt the plan. Now, however, the toddler is resolutely focused upon getting Toy A – he is unaware of the signals being sent by Nature to reevaluate the situation. The process
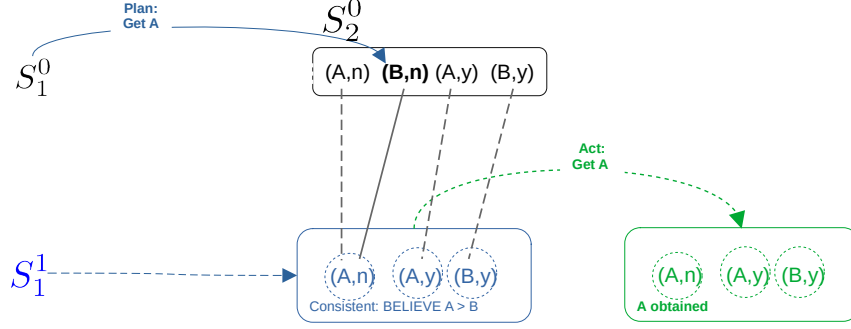
16

Figure 12: Individual 1's intended unawareness

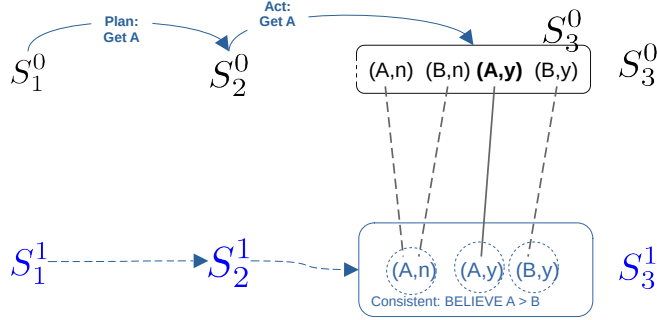is happily concluded as illustrated in Fig. 14.

Figure 13: Actualized state (B, n) does not disrupt the plan

An interesting connection

While thinking about this example, I came across the idea of OODA Loops (Observe, Orient, Decide, Act), which gained popular use in the military. The wiki discussion is here. I also downloaded a couple of short articles about this into the repo lit file. The diagram of the OODA Loop is shown in Fig.

The articles are interesting for at least two reasons. First, they are dealing with split-second
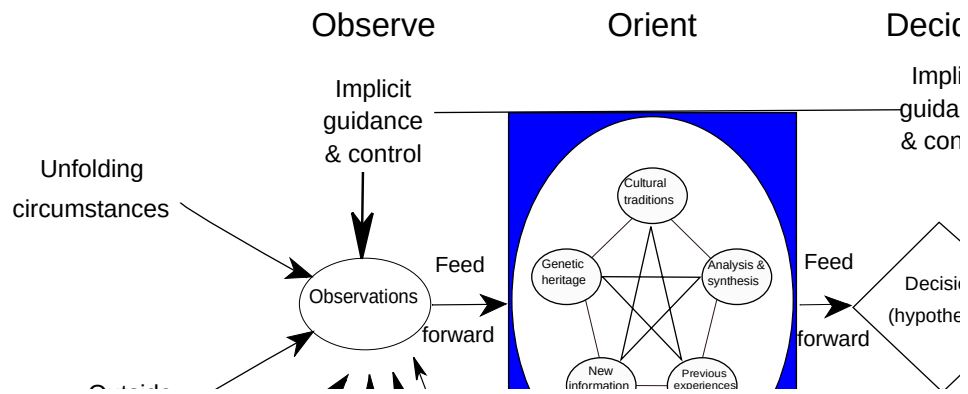
Figure 14: Mission accomplished

$$(A,A^*) \quad (A,B^*) \quad (B,A^*) \quad (B,B^*) \quad S_1^0$$

Figure 15: John Boyd's OODA Loop

decision situations, for example a soldier entering a building in hostile territory and having to decide whether to shoot at someone moving past an open doorway in the room ahead. The fact that this model is used for training people to make split-second decisions in life-or-death situations suggests there is something to it. Most importantly, it suggests that some process like this is going on at

the basic cognitive level and not only, e.g., at the level of higher-level, complex decisions that may involve explicit, more extended data gathering and analysis phases.

The other interesting angle is that this model assumes that interrupting the OODA loop resets the competitor's loop all the way back to the first stage. Hence, the explicit reason soldiers want to get inside the enemy's OODA loop is precisely that disrupting the process disrupts the enemy's ability to act. One way of thinking about our previous example is that Nature is "getting inside" the decision maker's "OODA loop".e

A deeper dive into the state spaces and FOAs in this example

The preceding discussion was intended as a first, rough cut exposition designed to outline and illustrate some key ideas. Having done that, there are some aspects worth refining.
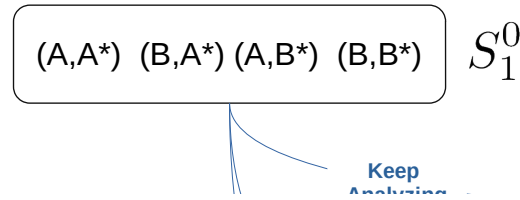

Figure 16: A more accurate set of Nature's states in $t = 1$

The previous elaboration of nature's state spaces was compressed in order to get at the overall framework without too much clutter. A more accurate represetnation is shown in Fig. 16. In $t = 1$, Toddler hears one of the Toys ringing a bell and believes one of the toys is best. The states are (x, y) where x is the toy that is ringing and y is the toy that is best. The toddler always believes that the ringing toy is going to be the best toy. Since acting is not allowed without a plan, there are not states in the first period in which a toy is obtained. Presumably, the truly best toy is determined by Nature at the start, in period $t = 1$.

In this example, the transition from S_1^0 to S_2^0 depends upon the act of the toddler in period 1. There are three possibilities: i) continue to gather information; ii) commit to a plan to get A; or iii) commit to a plan to get B. Each act leads to a corresponding Nature's state space ($S_{2,i}^0$ through $S_{2,iii}^0$ as illustrated in Fig. 17). In the second period, the actualized state (x, y, z)depends upon which toy is ringing in $t = 2$ (x), which toy is actually best (y) and, if operating under a

$S^0_{2,i}$ (A,A*) (B,A*) (A,B*) (B,B*)

Keep
Analyzing → (A,A*) (B,A*) (A,B*) (B,B*) $S^0_{3,i}$

Plan:
Get A → (A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D) $S^0_{3,ii}$

Plan:
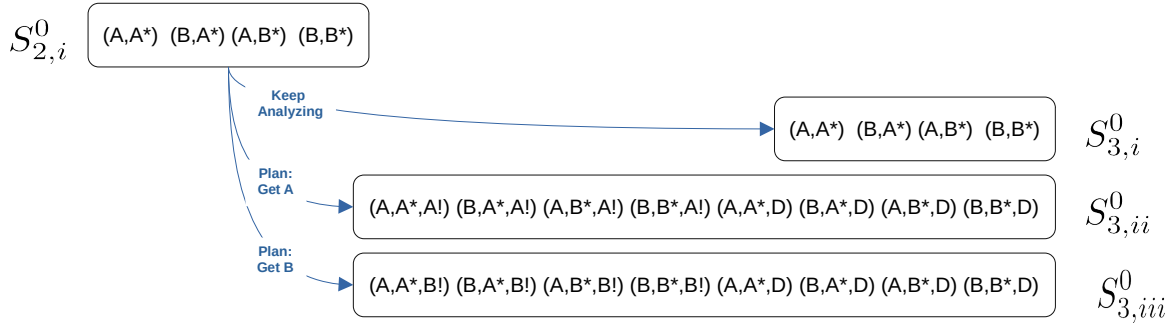Get B → (A,A*,B!) (B,A*,B!) (A,B*,B!) (B,B*,B!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D) $S^0_{3,iii}$

Figure 17: Possible state spaces in $t = 2$ depend upon toddler's act

committed plan, whether the state is consistent with the plan (e.g., $z = A!$ if the plan is to get A) or the toddler becomes aware of something that disrupts the plan ($z = D$).

The tree expands substantially in period $t = 3$. The possibilities are illustrated in Fig. 18. What happens in $t = 3$ depends upon the state space actualized in $t = 2$ in conjunction with the act of the toddler in $t = 2$. If the toddler continued analysis, then the possibilities are $S^0_{2,i}$ through $S^0_{2,iii}$, mirroring the possibilities in period $t = 2$. If the committed plan was to get A, then three possibilities are illustrated: Space $S^0_{3,iv}$, corresponding to the plan being disrupted in $t = 2$ and the toddler choosing to reanalyze the situation; Space $S^0_{3,v}$, corresponding to the plan being disrupted in $t = 2$ and the toddler committing to a plan to get B; and Space $S^0_{3,vi}$, corresponding to an actualized state consistent with the plan and the toddler acting to get A. The possibilities following $S^0_{2,iii}$ mirror these (shown as $S^0_{3,vii}$ through $S^0_{3,ix}$).

It is important to note that, although $S^0_{3,i}$, $S^0_{3,iv}$, and $S^0_{3,vii}$ appear to be identical (the analysis state space), in fact they are not. The reason for this is that each state also contains its own history. Since the histories leading to each of these state spaces is different, technically, the states they contain are not identical.

Finally, what do the toddler's FOAs look like in this example? Let us consider the sequence described in the previous, problematic case: Nature switches between ringing toys, starting with A, the toddler commits to the plan to get A, following which there is no disruption. As shown in Fig. 19, there are a couple of possibilities for the toddler's FOA that come immediately to mind. On the left-hand side, the toddler is aware of all the states but uncertain about which toy is best. His beliefs are such that he believes the ringing toy is best. On the right-hand side, the toddler
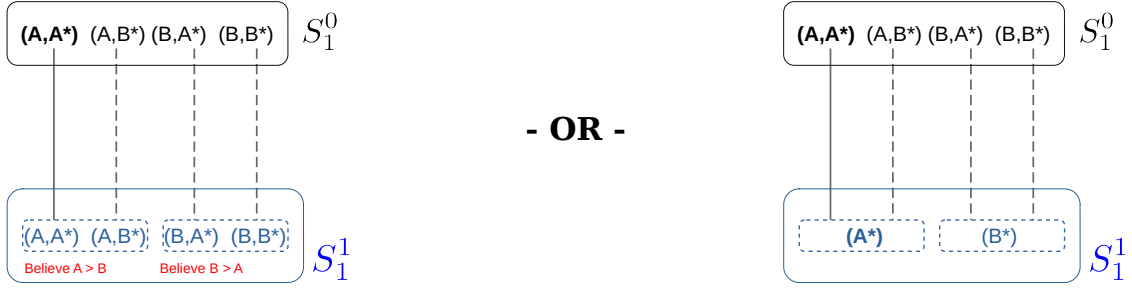
20

Figure 18: Possible state spaces in $t = 2$ depend upon toddler's act

is unaware of all the possibilities. Rather, he is certain that whichever toy is ringing is best. The latter is probably a good model of a toddler, the former would be better for a sophisticated decision maker.

In $t = 2$, illustrated in Fig. 20, the plan and its associated unawareness kick in. Even though toy B is ringing, the toddler's FOA projects all the plan-consistent states into a single, plan-to-get-A-consistent state, $(A!)$. Here, the toddler is shown as being aware of the possibility that something could happen to disrupt the state.

I am not sure this depiction is accurate. Since everything is going according to plan, including (D) may be incorrect. Rather, if something actually happened to disrupt the plan (Mom enters
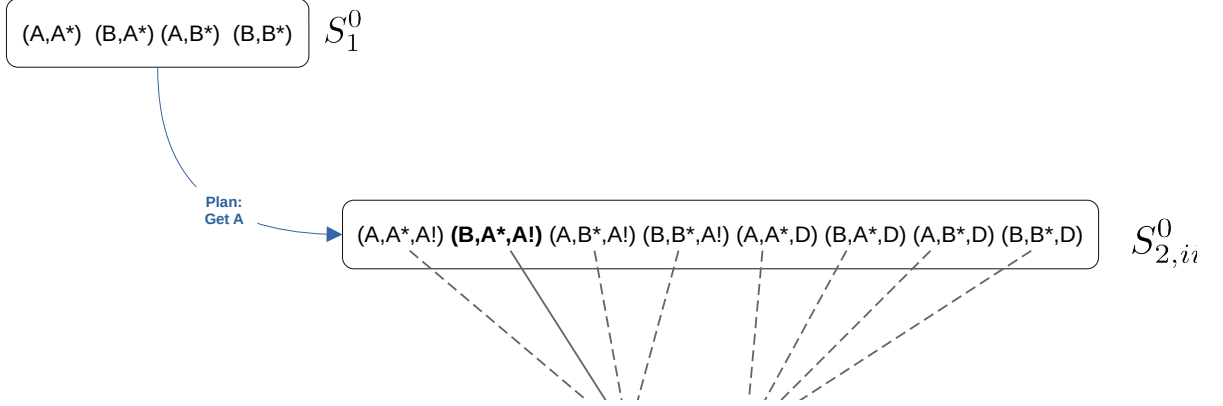
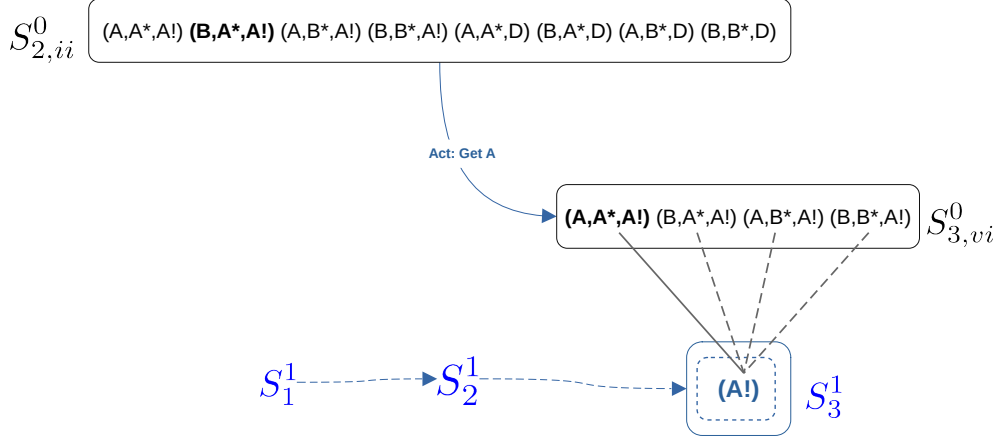Figure 19: Toddler believes the ringing toy is best

Figure 20: Plan coupled with intentional unawareness

the room and says, "It is time for bed, young man!"), then that would count as an act of Nature, resulting in the FOA shown (with the (D) state included in the FOA).

Finally, we come to the last stage, in which the toddler obtains A. This is shown in Fig. 21. This is fairly straightforward, though there are a couple of options here as well. In the top version, toddler gets A and is only aware of that – having achieved his intended end, he simply moves on to other things. Alternatively, the toddler may be aware of the true state (it really is the best toy!) and, as well, may be able to reason about the other possibilities. Note that all states in the toddler's FOAs have the $A!$ (get A) indicator, since this is accomplished by his act in $t = 2$.
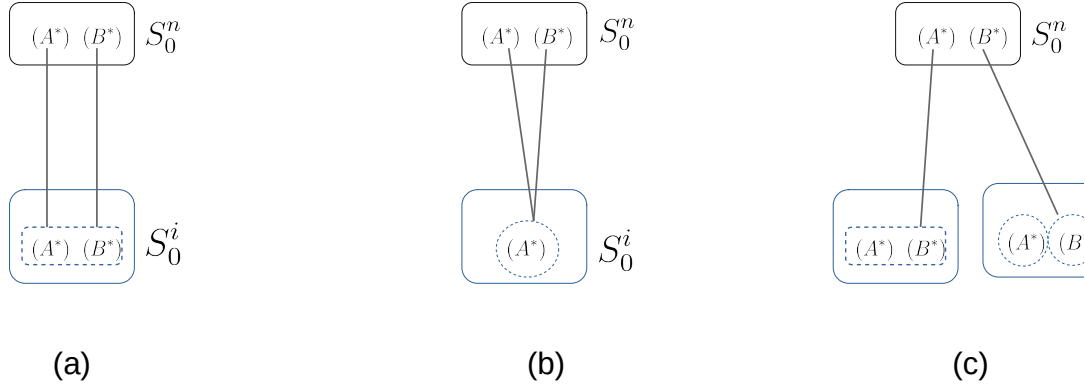
Figure 21: Final outcome possibilities: only aware of getting A or all relevant state details

Therefore, it is redundant (and could be removed). But, keeping the indicator there is fine since it is consistent with what happens in all states.

References