Examples of Awareness of Intentions[1]

Brian Epstein
Tufts University, Medford

Michael D. Ryall
University of Toronto

June 17, 2021

[1] Rough examples for the paper

We consider the problem of Brian's Toddler. The toddler, Individual 1, is faced with a decision as to which of two toys, labeled A and B, to obtain. The issue is whether A or B is best.

Parsimonious game theoretic treatment

The parsimonious game theoretic treatment 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.

A four-phase decision process

Our goal is to expand the parsimonious treatment of individual 1's cognitive process to include some of the features debated in the philosophy literature. With an eye toward adopting the unawareness formalism of game theory, we begin by elaborating what 1 is aware of about the world, how he thinks about his decision at a given moment in time, and how this evolves dynamically.

First, what is the decision process? One useful disaggregation of the decision process for our context is: **Phase 1** Individual finds himself in a state of the world in which a decision is called for (here, we count deciding not to pursue the decision further as, itself, a decision); **Phase 2** Individual selects focal elements for the decision analysis, analyses them, and forms an intention;

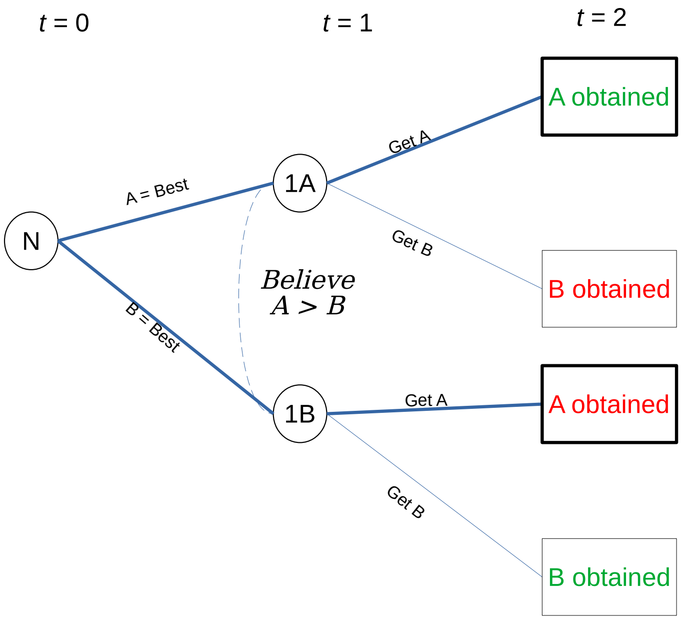

Figure 1: Brian's Toddler, parsimonious game-theoretic treatment**Phase 3**

Phase 3 Individual forms a plan to effectuate the intention; **Phase 4** Individual acts in accordance with the plan.

In our example: **Phase 1** Individual 1 is presented with a choice of A or B; **Phase 2** Individual 1 consults beliefs and forms an intention to obtain A; **Phase 3** Individual 1 plans to effectuate the intention by crawling to the location of Toy A and picking it up; **Phase 4** Individual 1 crawls to Toy A and picks it up.

Note that Phase 2 may involve limiting attention to a subset of what we might call Individual 1's "field-of-awareness" (FOA for short, which could also abbreviate field-of-attention if we wish). That is, given all the elements of the world of which 1 is passively aware at the start of the decision process, he may decide to limit his attention to a strict subset of elements thought to be decision-relevant. Similarly, the individual may call to mind (make himself aware) elements that expand his FOA. The plan formed in Phase 3 may be simple, but it may also be state-contingent (and, typically, will be for the satisfaction of complex real-world intentions). A key assumption is that the plan, once formulated, will be enacted as long as the world unfolds in a way that is "sufficiently consistent" with it (the precise meaning of this will need to be worked out).

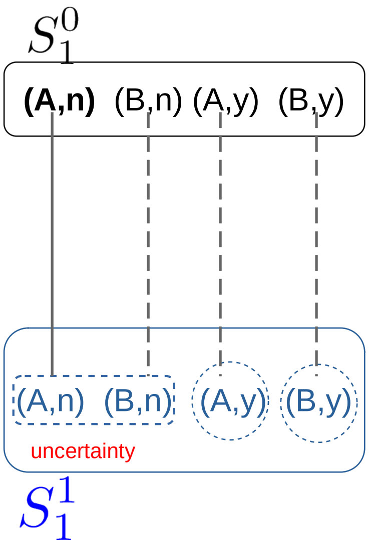

Figure 2: Toddler's Field-of-Awareness (blue)

Therefore, we begin by defining the states of the world at $t = 1$ as $S_1^0 \equiv \{(A, n), (B, n), (A, y), (B, y)\}$. This follows our earlier notation where S indicates a state space, 0 indicates Nature, and 1 indicates the time period. The states are (x_1, x_2) where x_1 is which toy is truly best and x_2 is whether a toy is obtained or not (note: for completeness, the state should elaborate *which* toy is obtained, but the example will work without the extra clutter).

Begin with the case of a fully aware decision maker. Assume the state is (A, n) (indicated by the bold typeface). In this state, Individual 1's FOA is $S_1^1 = S_1^0$ as illustrated in Fig. 2 by the blue part of the diagram. The other states in S_1^0 are true counterfactuals – they are states that really could have been actualized in $t = 1$ given some other sequence of historical events.

The solid grey line shows that the FOA depicted is the one that arises in state (A, n). This is important: states encode everything about reality in that slice of time – this includes the FOA of the toddler. We could make things more explicit by writing $S_1^1(A, n)$; i.e., the FOA is a function of the actualized state. Since this should be unambiguous from the example, we omit the additional notation. Nevertheless, each of the four states, ostensibly, result in different FOAs. Since the other states are counterfactuals, the FOAs with which they are associated are omitted.

The grey dashed lines project the true counterfactual states of the world to their counterparts in the decision maker's actualized FOA. These projections from states in S_1^0 to S_1^1 indicate how counterfactuals in reality map to those in the decision maker's FOA. For example, the toddler is able to reason about the counterfactual state "(B, n)" in his FOA and, when he does, the state he is reasoning about corresponds to Nature's true state (B, n).

The blue dashed lines indicate the toddlers *information sets*. These are collections of states in the toddler's FOA that he believes *could be* true. The dashed circles around (A, y) and (B, y) indicate that the toddler is aware that were, e.g., (A, y) true he would know it. In a game theory model, the toddler would typically be endowed with beliefs in the form of a subjective probability distribution on states within the information set (the figure simply indicates uncertainty exists – we will be more specific momentarily).

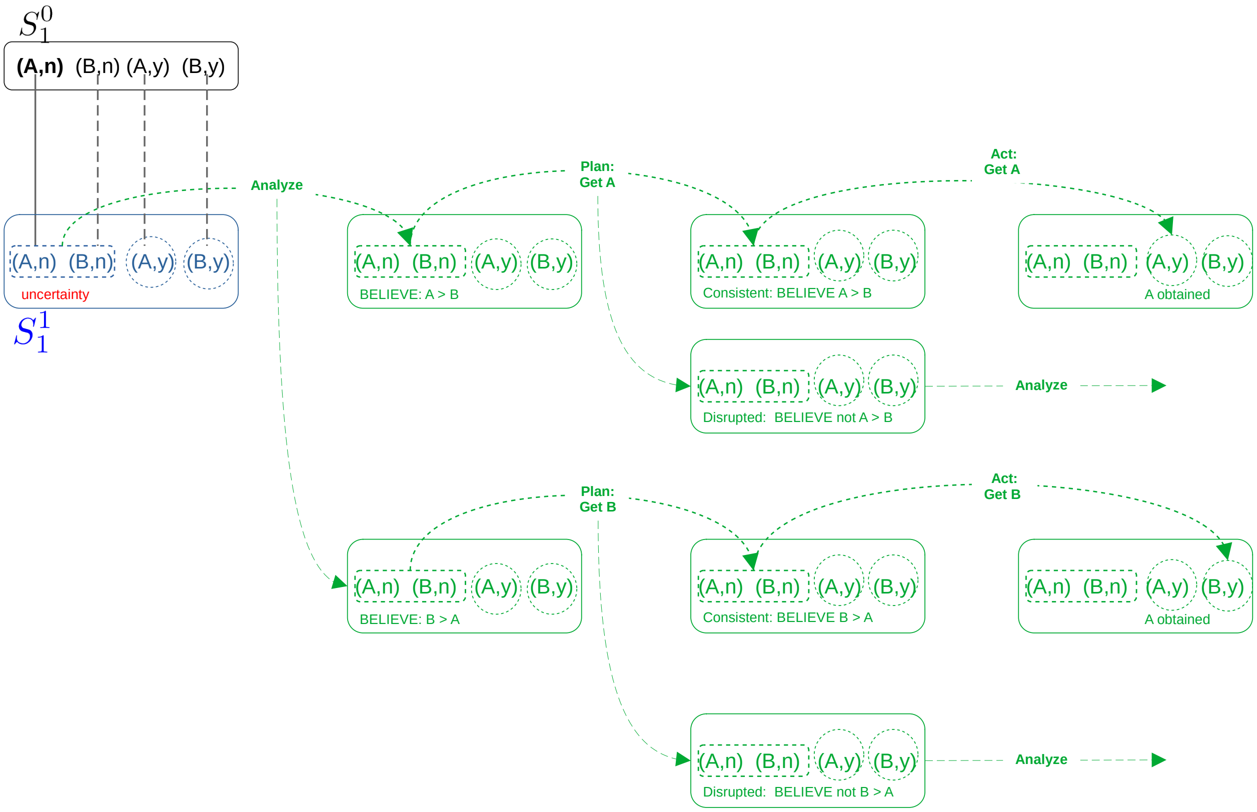

Figure 3: At $t = 1$, Toddler thinks about the future (green)

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 in Fig. 3. Moving to Phase 2, which is an analysis and commitment step, 1 may discover that either A or B is preferred. Here, analysis takes one time period. Once the analysis is complete, and the intention is formed to obtain the preferred toy, 1 formulates a plan. The planning takes one time period.

Here, we make an important assumption: while the individual is thinking, analyzing, planning, and acting, the world evolves. Now, from $t = 1$ to $t = 2$, this evolution is implicit as occurring and possibly affecting the analysis. The individual considers things from the perspective of his FOA, which may evolve during the interval between periods. Still, the outcome of that process is that 1 comes to believe either that A or B is best. What happens from $t = 2$ to $t = 3$ is distinct from the analysis phase. During this interval, a plan to get A or to get B is being shaped. However, real-world events may intrude upon the process in ways that disrupt the plan. This is illustrated in the figure. For example, 1 may be planning to get A yet experience something that happens to indicate that B is really best (e.g., Toy B starts making a ringing sound). When the plan is disrupted, we assume that the decision maker must backtrack to an earlier stage – either a new analysis or a new planning cycle. In the figure, we show that the process reverts to a new analysis stage.

Then, if the state of the world in $t = 3$ is consistent with the plan plan, individual 1 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. In $t = 4$, therefore, the final state is (A, y) (and 1 knows that this is the state).

A two-phase reduction

Before continuing with the example let us compress Phases 1-3 into one – i.e., there is a thinking, deciding, and planning phase followed by an acting phase. This seems to give us a sufficient level of elaboration to investigate the kinds of issues in which we are interested. Refer to the compressed phase as "planning." The revised version of Fig. 3 is shown in Fig. 4.

All of this is by way of setting up the dynamic analysis. Based upon the, now, compressed analysis, decision, and planning phase, 1 develops a plan to get A. Fig 5 shows the situation at the start of $t = 2$. The world has evolved to S_2^0 in which state (A, n) continues to hold. Reality is illustrated on the top row. Individual 1's act "Plan Get A" happens and leads to S_1^0.

Keep in mind that the state labelled (A, n) in S_2^0 is not identical to the one so labelled in S_1^0, even though the state spaces appear to be the same. First, the states in S_2^0 includes the information

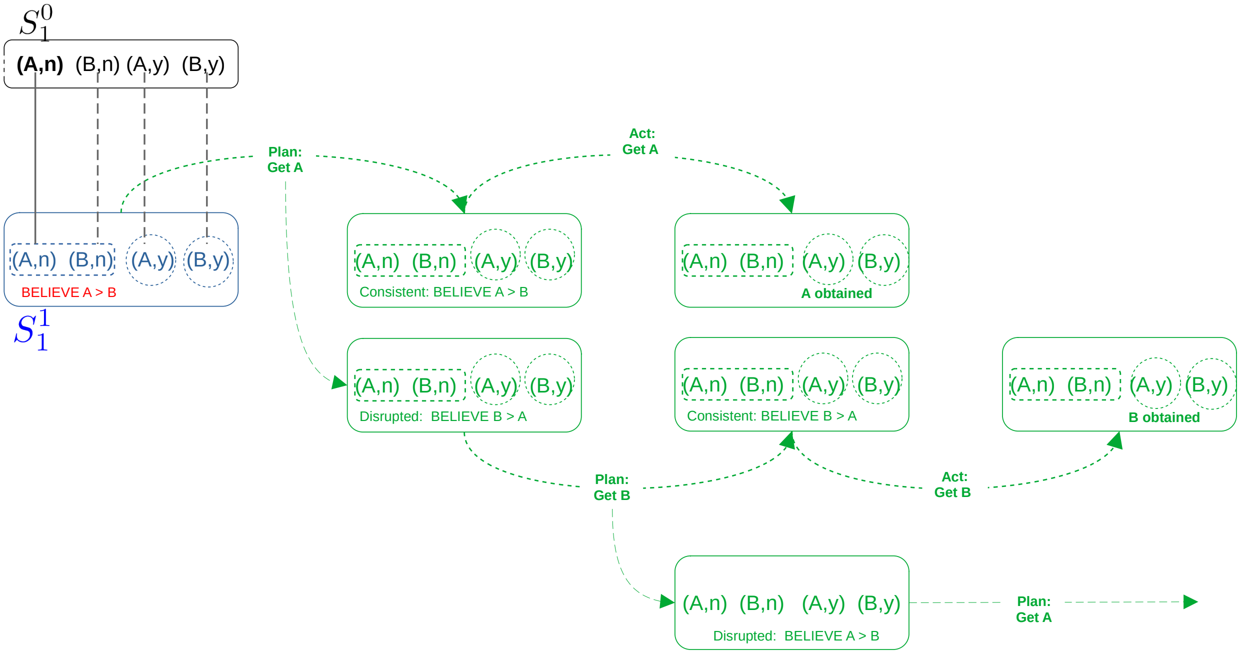

Figure 4: The two-phase version

about the sequence of preceding states (i.e., $s_1^0 = (A, n)$) as well as about the acts that caused them (i.e., a_1^1 = plan get A). 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_1^1 as well as what action he took. 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.

Having formed a plan to get A, nothing has happened to disrupt 1's decision to get A. His belief remains that A is best. As we saw above, an alternative possibility was that an act of Nature, e.g., Toy B begins ringing, might have disrupted the plan. This emphasizes that what we are illustrating is one path of actual events analogous to the bold lines through the tree in Fig. 1. A diagram of all possible paths and FOAs would be quite complex to illustrate in a single diagram.

Evolution to the final period, $t = 3$ is shown in Fig. 6. At the conclusion, individual 1 obtains A and is certain that he obtained A.

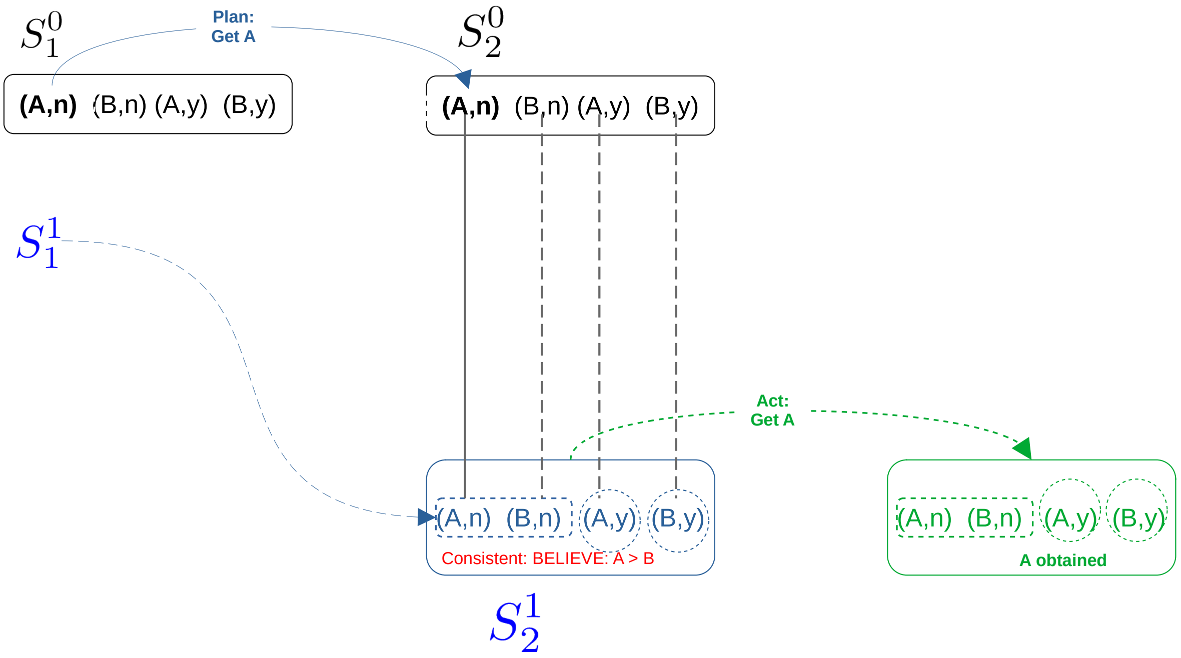

Figure 5: The plan proceeds without disruption

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 1 continues to have full awareness and, as well, his projection into the future is consistent with reality. One would need to be careful to ensure that feasible actions available to 1 would correspond to the logic of our assumptions. In particular, acts must follow plans and, presumably, some states could disrupt the plan which would leave 1 with one feasible act (restarting the planning process) or, if we include quitting the decision, two feasible acts. That game would be the one shown in Fig. 1 extended to give nature intervening moves (i.e., with the power to switch 1's assessment of which is the better toy) and adjusting the feasible acts as described.

Our diagrams give a more elaborate description of what is going on in 1's mind (recalled history, FOA, and projected futures). However, this comes at the expense of providing a simultaneous illustration of the entire set of possible paths as we have in extensive form game trees. Later, if we introduce unawareness, our FOAs will provide an expressive capacity not available to standard

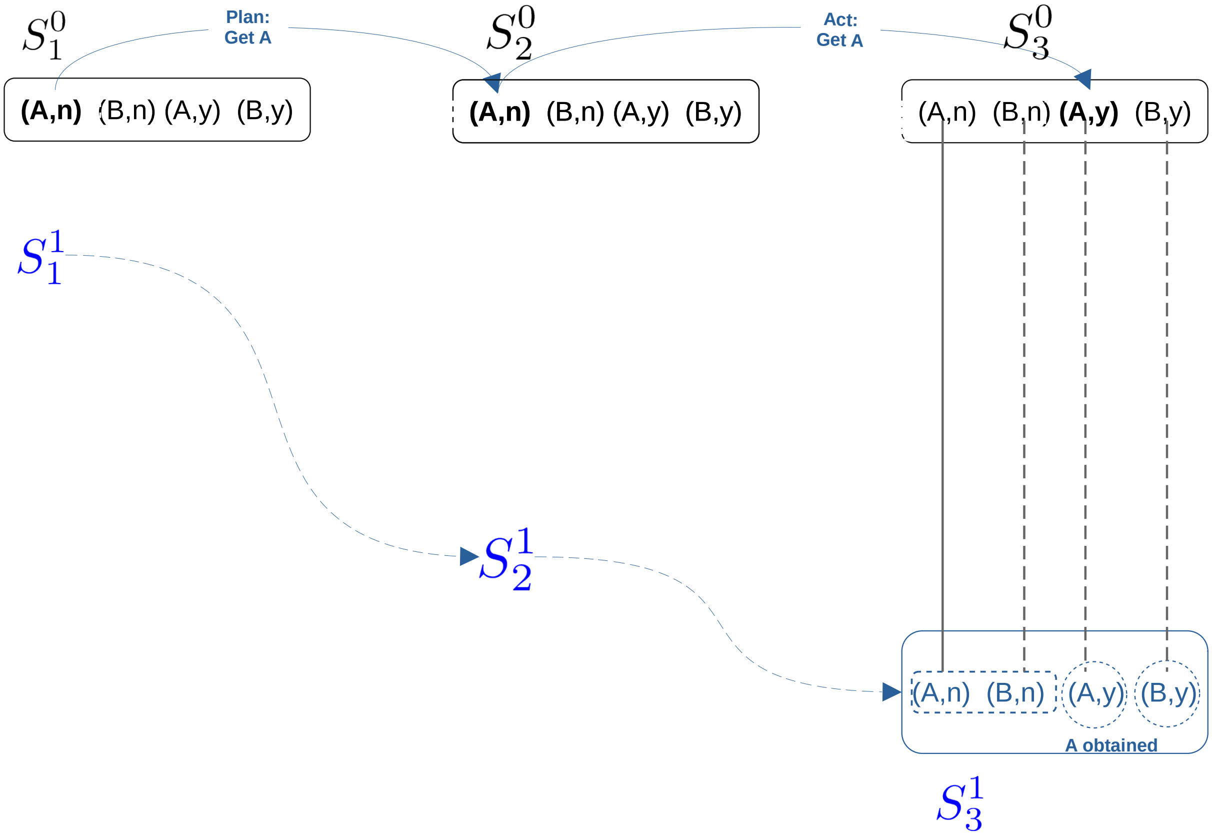

Figure 6: Decision and action resolve

extensive form games.

The next step is to illustrate what can go wrong – how an even fully aware individual (or, perhaps, a fully aware individual in particular) can get stuck in a "paralysis by analysis" situation. Then, we can show why intentional unawareness can help (or hurt) the situation.

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. 7: 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.

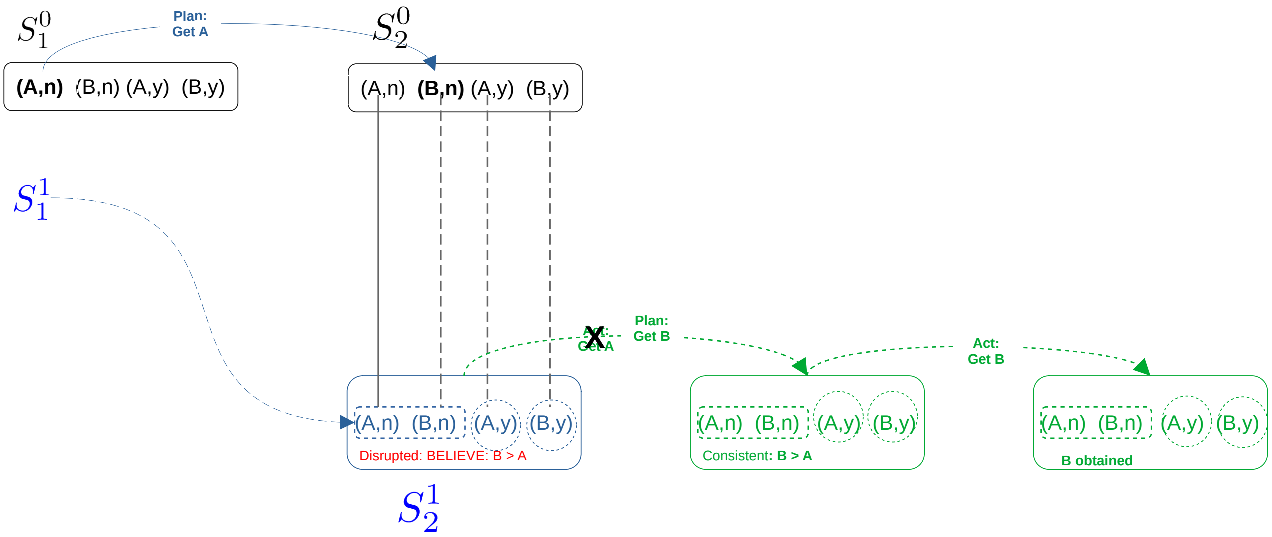

Figure 7: Nature disrupts the plan

This alternating flow of information can go on indefinitely, as illustrated in Fig. 8. Under the standard, Bayesian belief-desire model, a rational agent is one who immediately updates beliefs 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.)

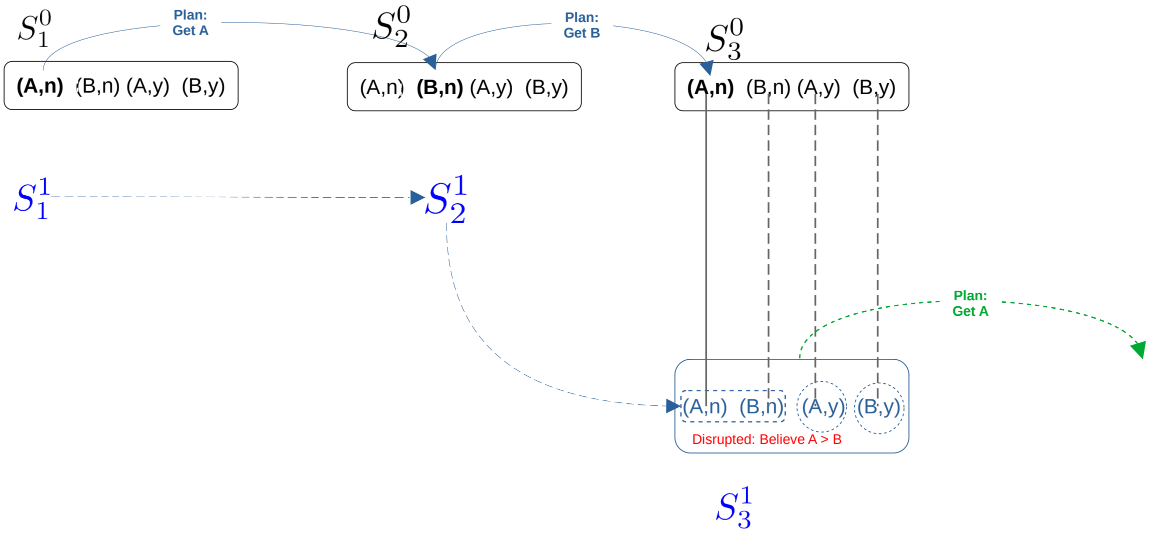

Figure 8: Nature disrupts the plan again

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 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. 9 illustrates the situation in $t = 1$ in which the intention, plan, and planned unawareness all occur in the period.

Notice the change from Fig. 4: 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.

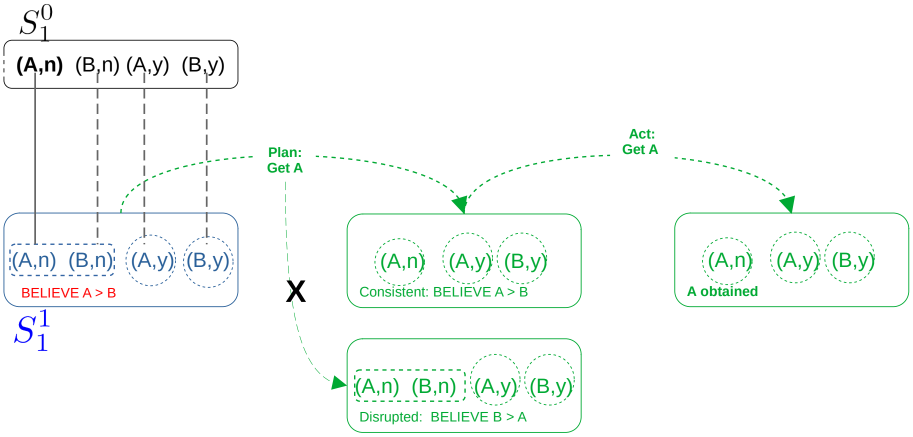

Figure 9: Individual 1's intended unawareness

The implementation of the plan is shown in Fig. 10. 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 is happily concluded as illustrated in Fig. 11.

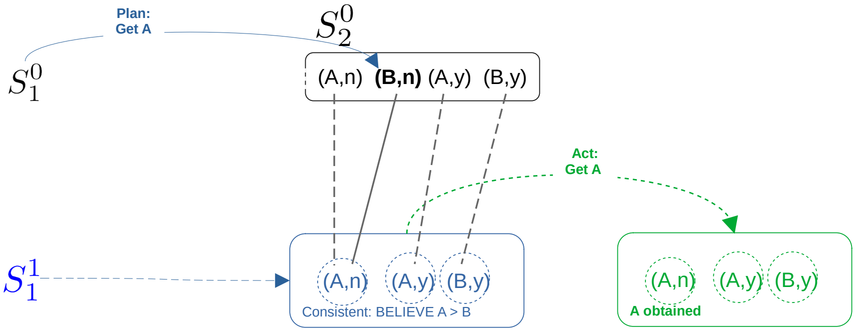

Figure 10: Actualized state (B, n) does not disrupt the plan

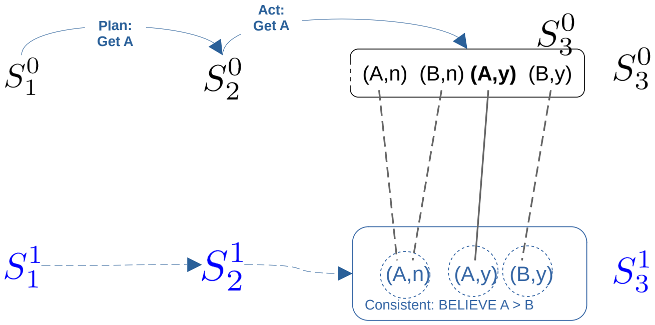

Figure 11: Mission accomplished

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 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

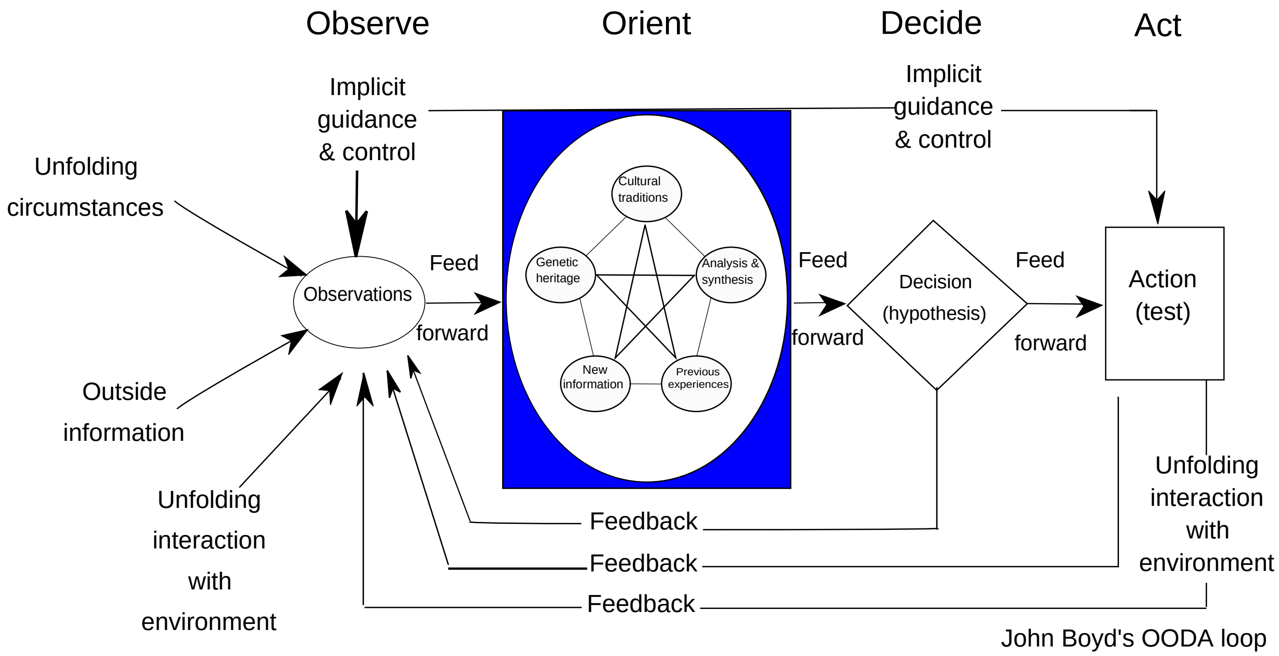

Figure 12: John Boyd's OODA Loop

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.

(A,A*) (A,B*) (B,A*) (B,B*) S_1^0

Figure 13: 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. 13. 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 $t = 1$.

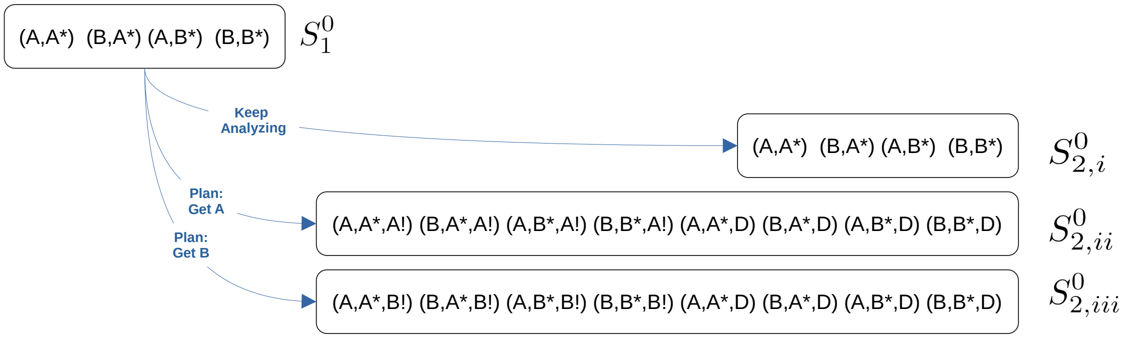
(A,A*) (B,A*) (A,B*) (B,B*)
S_1^0
Keep Analyzing
(A,A*) (B,A*) (A,B*) (B,B*)
$S_{2,i}^0$
Plan: Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S_{2,ii}^0$
Plan: Get B
(A,A*,B!) (B,A*,B!) (A,B*,B!) (B,B*,B!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S_{2,iii}^0$

Figure 14: Possible state spaces in $t = 2$ depend upon toddler's act

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. 14). 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 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. 15. 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 $S_{2,i}^0$ through $S_{2,iii}^0$, mirroring the possibilities in period $t = 2$. If the committed plan was to get A, then three possibilities are illustrated: Space $S_{3,iv}^0$, corresponding to the plan being disrupted in $t = 2$ and the toddler choosing to reanalyze the situation; Space $S_{3,v}^0$, corresponding to the plan being disrupted in $t = 2$ and the toddler committing to a plan to get B; and Space $S_{3,vi}^0$, corresponding to an actualized state consistent with the plan and the toddler acting to get A. The possibilities following $S_{2,iii}^0$ mirror these (shown as $S_{3,vii}^0$ through $S_{3,ix}^0$).

It is important to note that, although $S_{3,i}^0$, $S_{3,iv}^0$, and $S_{3,vii}^0$ appear to be identical (the analysis

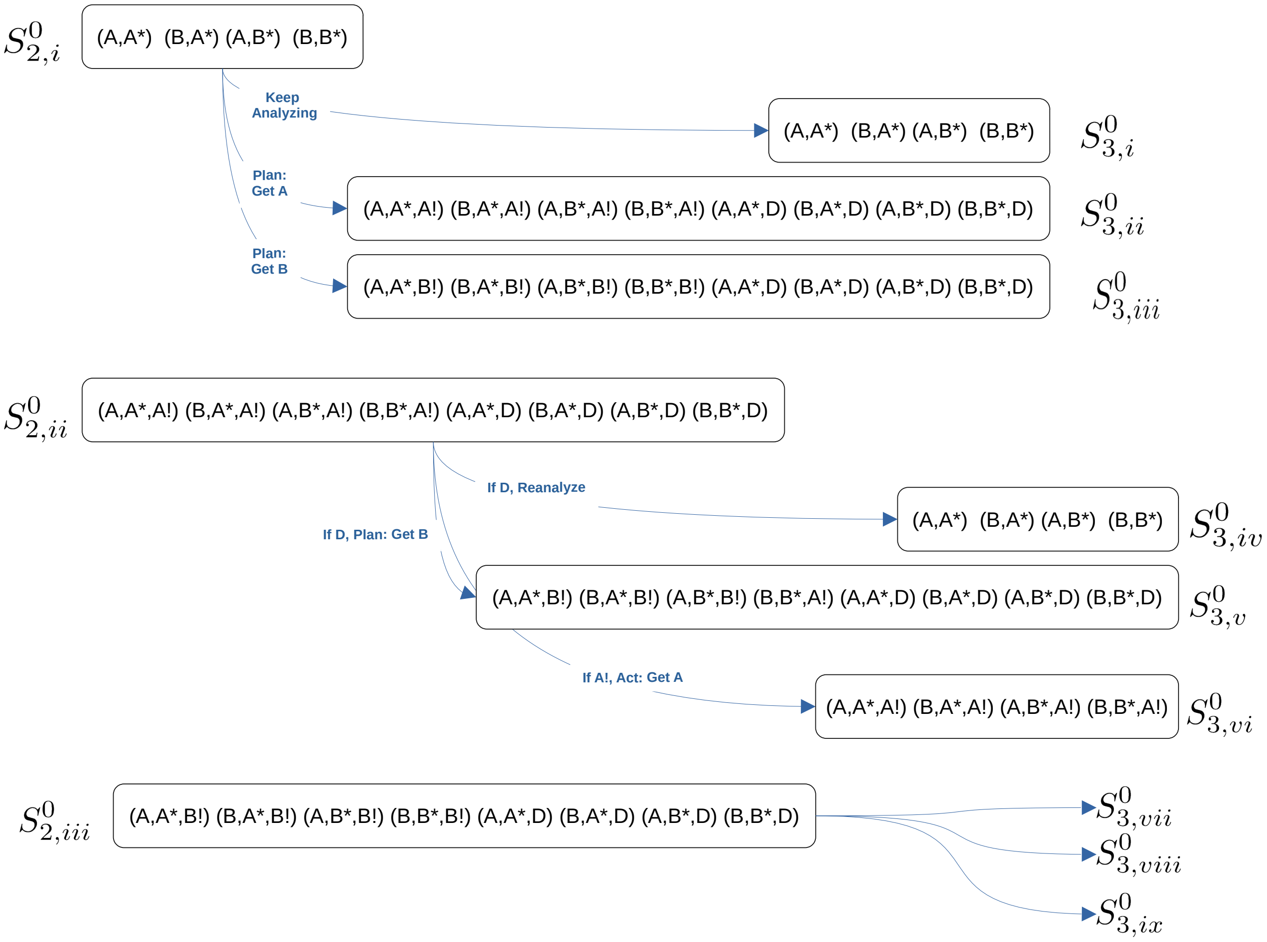
$S^0_{2,i}$
(A,A*) (B,A*) (A,B*) (B,B*)
Keep Analyzing
(A,A*) (B,A*) (A,B*) (B,B*)
$S^0_{3,i}$
Plan: Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S^0_{3,ii}$
Plan: Get B
(A,A*,B!) (B,A*,B!) (A,B*,B!) (B,B*,B!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S^0_{3,iii}$
$S^0_{2,ii}$
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
If D, Reanalyze
(A,A*) (B,A*) (A,B*) (B,B*)
$S^0_{3,iv}$
If D, Plan: Get B
(A,A*,B!) (B,A*,B!) (A,B*,B!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S^0_{3,v}$
If A!, Act: Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!)
$S^0_{3,vi}$
$S^0_{2,iii}$
(A,A*,B!) (B,A*,B!) (A,B*,B!) (B,B*,B!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S^0_{3,vii}$
$S^0_{3,viii}$
$S^0_{3,ix}$

Figure 15: Possible state spaces in $t = 2$ depend upon toddler's act

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. 16, 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

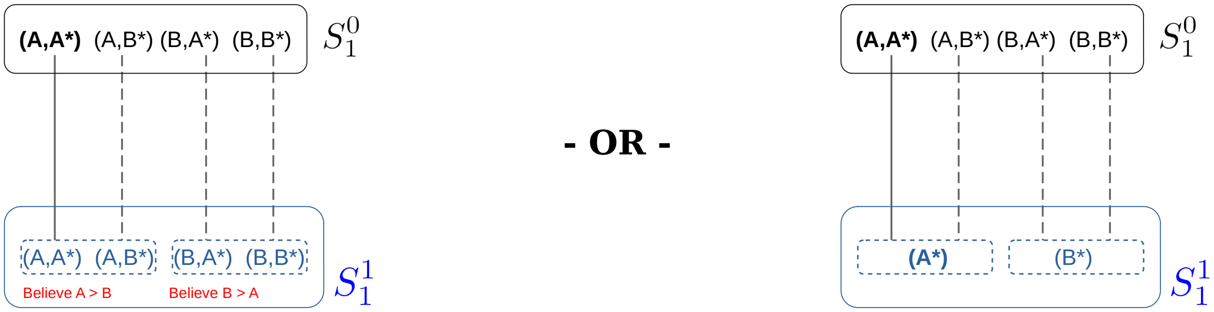

Figure 16: Toddler believes the ringing toy is best

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.

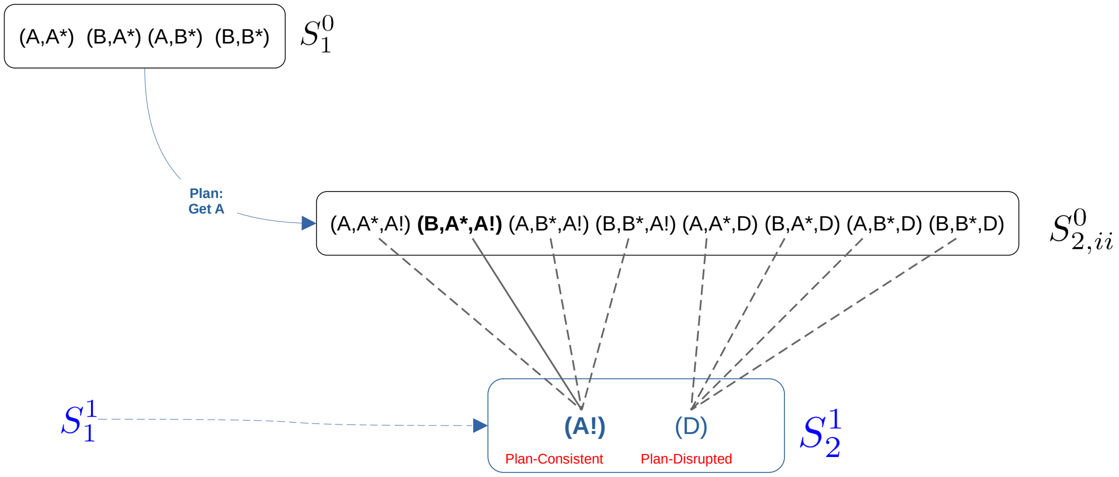
(A,A*) (B,A*) (A,B*) (B,B*)
S_1^0
Plan:
Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
$S_{2,ii}^0$
S_1^1
(A!)
(D)
Plan-Consistent
Plan-Disrupted
S_2^1

Figure 17: Plan coupled with intentional unawareness

In $t = 2$, illustrated in Fig. 17, 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 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).

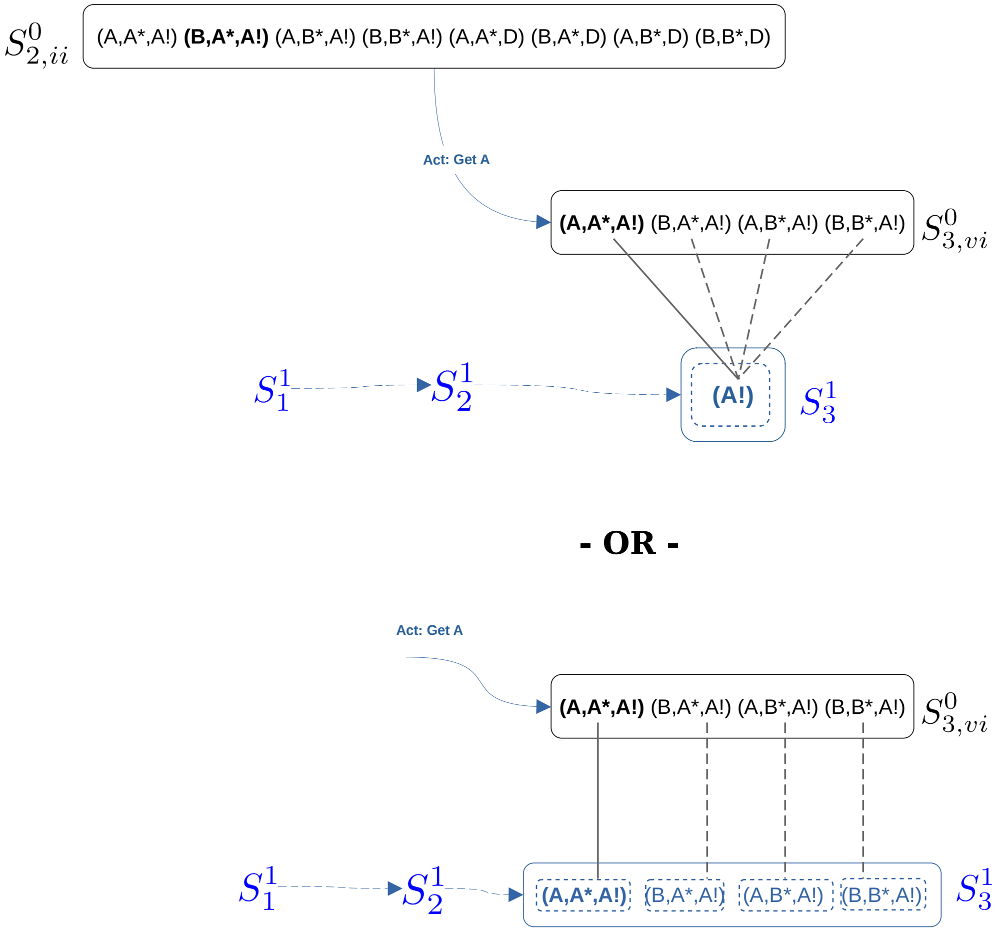
$S^0_{2,ii}$
(A,A*,A!) **(B,A*,A!)** (A,B*,A!) (B,B*,A!) (A,A*,D) (B,A*,D) (A,B*,D) (B,B*,D)
Act: Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!)
$S^0_{3,vi}$
S^1_1
S^1_2
(A!)
S^1_3
- OR -
Act: Get A
(A,A*,A!) (B,A*,A!) (A,B*,A!) (B,B*,A!)
$S^0_{3,vi}$
S^1_1
S^1_2
(A,A*,A!)
(B,A*,A!)
(A,B*,A!)
(B,B*,A!)
S^1_3

Figure 18: Final outcome possibilities: only aware of getting A or all relevant state details

Finally, we come to the last stage, in which the toddler obtains A. This is shown in Fig. 18. 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 $t = 2$.

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