Breaking Down Goal-Directed Behaviour

Breaking Down Goal-Directed Behaviour

Crossposted from the AI Alignment Forum. May contain more technical jargon than usual.

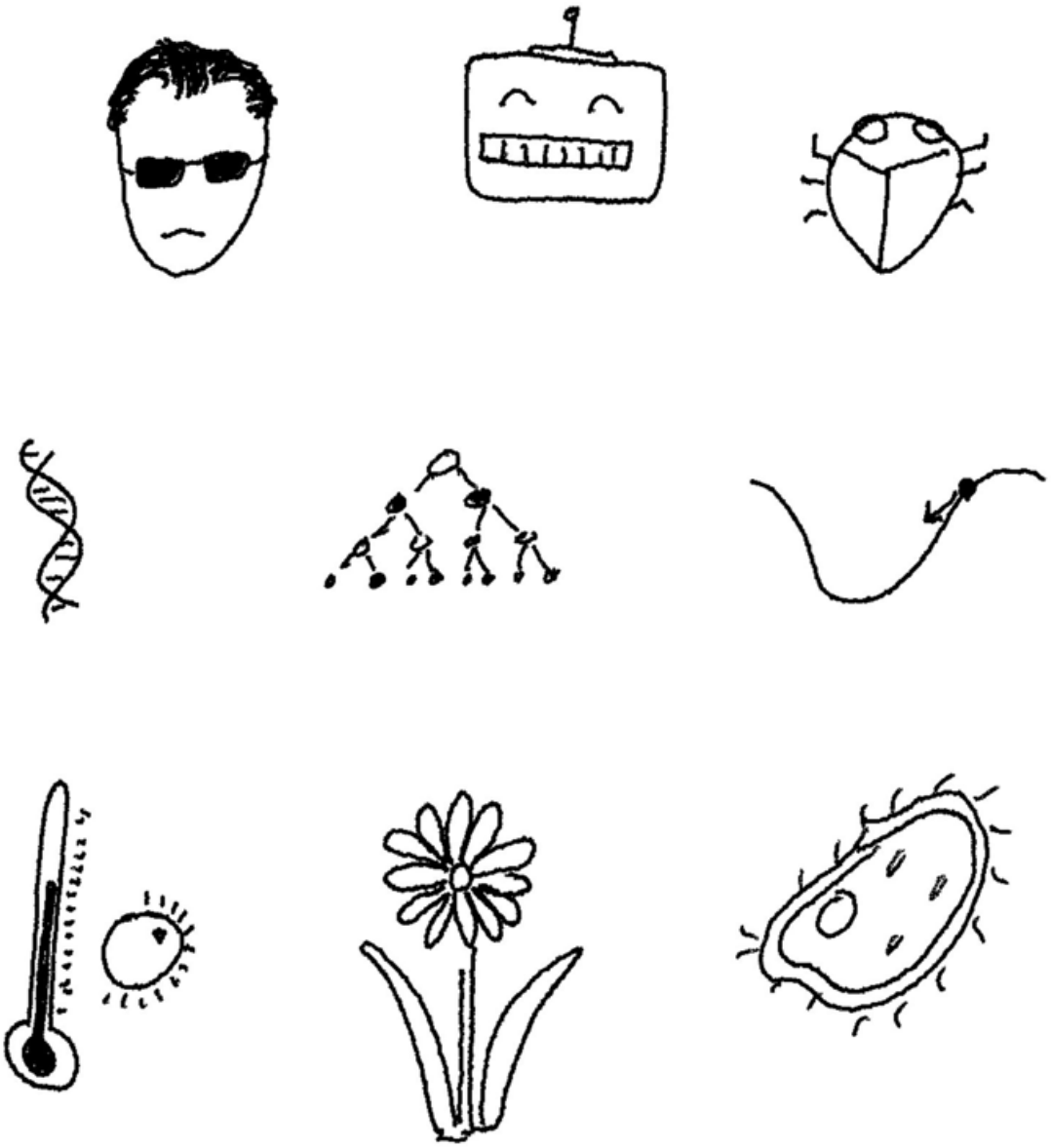
When we speak about entities 'wanting' things, or having 'goal-directed behaviour', what do we mean?

Because most of the actors that we (my human readers and I) attentively interact frequently with are (presumably) computationally similar (to each other and to ourselves), it is easy for our abstractions to *conflate* phenomena which are in fact different[1], or to *couple together* impressions of phenomena which are in fact separable[2]. On the occasions that we attentively observe *dissimilar* actors[3], we are often-enough 'in the wild' (i.e. looking in their 'habitat') that conflating - anthropomorphising - is sufficiently usefully predictive to get by[4], so these poor abstractions are insufficiently challenged.

Further, most of the people who in fact attentively observe a particular class of human-*dissimilar* actors (enough to perceive and understand the non-conflations) are focused domain experts about that particular class, and talk mainly to other domain experts about them[5]. There are few venues in which it is useful to have unambiguous terminology here. Thus, the language we use to *communicate* our abstractions about goal-directed behaviour is prone to conflation and confusion even when *some* people have *some* of the right abstractions[6].

Here I aim to take steps to break down 'goal-directed behaviour' into a conceptual framework of computational abstractions for which I offer tentative terminology, and which helps me to better understand and describe analogies and disanalogies between various goal-directed systems. The overarching motivation is to better understand goal-directed behaviour, in the sense of being able to better predict its (especially counterfactual and off-distribution) implications, its arisal, and other properties. Hopefully it is clear why I consider this worthwhile.

In order to ground this discussion, I refer to a reasonably diverse menagerie of candidate goal-directed systems, including natural and artificial systems at various levels of organisation. Contemplation of this diverse collection was responsible for the ideation and refinement of the ideas and gives some confidence in the appropriateness of the abstractions.



A collection of a few 'agents' drawn from the menagerie

---

1. Different in the sense that, even if the observed surface phenomena are similar 'in the wild', their behaviour in different contexts might radically come apart; that is, a conflation is a poor predictor for out-of-distribution behaviour. ↩

2. Separable meaning that the absence of one or other piece is a meaningful, conceivable state of affairs, even if in practice they are almost always found composed together; that is, a coupled abstraction means if we have an impression of the one, we (perhaps incorrectly) assume presence of the other(s). ↩

3. That is, dissimilar from humans and from each other e.g. ants, genes, chess-engines, learning algorithms, corporations... ↩

4. If it is not clear why 'in the wild' (or 'in the typical setting') is important for the predictiveness of the anthropomorphism-heuristic, hopefully **Deliberation and Reflexes** and **Where does Deliberation Come From?** will clarify. In short, an actor fit for a particular setting can carry out 'deliberate-looking' behaviours without 'deliberative machinery', because the *process which generated* the actor provides enough (slow, gradual) 'deliberation' to locate such behaviours and bake them into 'reflexes'. ↩

5. e.g. entomologists, ornithologists, AI researchers, business executives, economists... are not often in a room together, at least not mutually-knowingly in their capacity as said experts of their respective fields ↩

6. I attempt to avoid use of the term 'agent' as it is a very loaded term which carries many connotations. In fact it is unfortunately a perfect exemplary linguistic victim of the abstractive conflation and coupling phenomena I have described. (I think the recent reception of the Gato paper was confused in part as a consequence of this.) I substitute 'actor' and 'controller' more freely as less loaded terms. ↩

You Only Get One Shot: an Intuition Pump for Embedded Agency

Crossposted from the AI Alignment Forum. May contain more technical jargon than usual.

This is a short attempt to articulate a framing which I sometimes find useful for thinking about embedded agency. I noticed that I wanted to refer to it a few times in conversations and other writings.

A useful stance for thinking about embedded agents takes as more primitive, or fundamental, 'actor-moments' rather than (temporally-extended) 'agents' or 'actors'. The key property of these actor-moments is that they get one action - one opportunity to 'do something' - before becoming simply part of the history of the world: no longer actual.

This is just one of the implications of embedded agency, but sometimes pulling out more specific consequences helps to motivate progress on ideas. It is an intuition pump, and, as with the exemplar archetype for intuition pumps, it does not tell the whole story and should be used with caution.

The Cartesian picture

It is often convenient to consider a decision algorithm to persist through time, separated from its environment by a Cartesian boundary. The agent receives (perhaps partial) observations from the environment, performs some computation, and takes some action (perhaps updating some internal state, learning from observations as it goes). The resulting change in the environment produces some new observation and the process continues.

This is convenient because it is often empirically approximately true for the actors we encounter[1].

Only one shot

In contrast, in reality, any act, and the concomitant changes in the environment, impinge on the actor (which is after all part of the environment), even if only in a minor way[2].

Taking an alternative stance where we imagine an actor *only existing for a moment* - having a single 'one shot' at action - can prompt new insights. In this framing, a 'state update' is just a special case of the more general perspective of 'self modification', which is itself a special case of 'successor engineering'. And all of these are part of the larger picture implied by the stance taking as more fundamental not 'actor' but 'actor-moment': wherein an actor makes a decision and so unfolds a particular future world trajectory, which may - or may not - contain relevantly-similar actor-moments.[3]

This stance is somewhat unnatural for many actors we encounter (and consider worthy of attention and conversation investment), because for the most part these

have been selected for a semblance of self-integrity and durability, and so the Cartesian abstraction is a workable approximation. But departures from these modes, or attraction to additional modes of influence over subsequent actor-moments, may be more accessible to future agents, especially artificial ones[4].

Further, I have found this a useful stance for analysing existing and hypothetical systems, because it can prevent leaking intutions, and it provides an alternate framing to better understand systems-building-systems (some which already exist, and some which might some day exist).

Finally, as embedded agents ourselves, humans can sometimes get useful insights from taking this stance.

Related

Scott Garrabrant discusses Cartesian Frames which I think are a related (and more fleshed out) conceptual tool.

---

1. most likely as a consequence of self- and goal-content preservation being simultaneously instrumentally and intrinsically useful strategies with respect to natural selection. ↩

2. Sometimes we can abstract that impact as consisting of a 'state update' to some privileged computational component of the algorithm implemented by the actor, while leaving the rest of the algorithm unchanged, and be essentially correct. ↩

3. All of this is ignoring the challenge of defining what an 'act' is, what an 'actor' is, and the related questions about counterfactuals. I also ignore the challenge of identifying the time interval constituting a 'moment': it seems sensible for this to depend on the essential form of the algorithm the actor implements. ↩

4. For example by uncoupling from constraints which restrict operations relevant to other instrumental goals ↩

Deliberation, Reactions, and Control: Tentative Definitions and a Restatement of Instrumental Convergence

Crossposted from the AI Alignment Forum. May contain more technical jargon than usual.

This analysis is speculative. The framing has been refined in conversation and private reflection and research. To some extent it feels vacuous, but at least valuable for further research and communication.

A cluster of questions fundamental to many concerns around risks from artificial systems regard the concepts of search, planning, and 'deliberateness'. How do these arise? What can we predict about their occurrence and their consequences? How strong are they? What *are* they anyway?

Here is laid out one part of a conceptual decomposition which maps well onto many known systems and may allow further work towards answering more of those questions. The ambition is to really get at the heart of what is algorithmically happening in 'optimising systems', including humans, animals, algorithmic optimisers like SGD, and contemporary and future computational artefacts. That said, I do not have any privileged insight into the source code (or its proper interpretation!) for the examples discussed, so while this framing has already generated new insights for me, it may or may not be 'the actual algorithmic truth'.

We start with the analysis: a definition of 'deliberation' and its components, then of 'reactions' and 'control'. Next we consider, in light of these, what makes a deliberator or controller 'good'. We find conceptual connections with discussions of instrumental convergence. Little attention is given here to *how to determine* what the goals are, which is obviously also important.

These concepts were generated by contemplating various aspects of many different goal-directed systems and pulling out commonalities.

Some readers may prefer to start with the examples, which include animals, plants, natural selection, gradient descent, bureaucracies, and others. Here in the conceptual section I'll footnote particularly relevant concrete examples where I anticipate them helping to convey my point.

A full treatment is absent, but two major deferences to embedded agency underlie this analysis. A Cartesian separation need not be assumed, except over 'actor-moments' rather than temporally-extended 'actors'. And a major driver for this sequence is a fundamental recognition that any goal-directed behaviour instantiated in the real world must have bounded computational capacity per time[1].

Inspiration and related

My (very brief) 'Only One Shot' intuition pump for embedded agency may help to convey some background assumptions (especially regarding how time and actor-moments fit into this picture).

Scott Garrabrant's (A → B) → A talks about 'agency' and 'doing things on purpose'. I'm trying to unpack that further. The (open) question Does Agent-like Behaviour Imply Agent-like Architecture? is related and I hope for the perspective here to be useful toward answering that question.

Alex Flint's excellent piece The ground of optimization informs some of the perspective here, especially a focus on *scope of generalisation* and *robustness to perturbation*.

Daniel Filan's Bottle Caps Aren't Optimisers and Abram Demski's Selection vs Control begin discussing the *algorithmic internals* of optimising systems, which is the intent here also. Risks from Learned Optimization is of course relevant.

John Wentworth's discussions of *abstraction* (for example What is Abstraction?) especially with regards its predictive properties for a non-omniscient computer, are central to the notion of abstraction employed here. Related are the good and gooder regulator theorems, which touch closely computationally upstream of the aspects discussed here while making fewer concessions to embedded agency.

Definitions

Deliberation

A *deliberation* is any part of a decision algorithm which composes the following

- Propose: generate candidate proposals
- Promote: promote and demote proposals according to some criterion
- Act: take outcome of promotion and demotion to activity in the environment

taking place over one 'moment' (as determined by the particular algorithmic embodiment).

Some instances may fuse some of these components together.

Propose

A deliberation needs to be *about something*. Candidate proposals X do not come from nowhere; some state S is mapped to a distribution over nonempty proposal-collections (there must be at least one candidate proposal[2]). Proposals correspond in some way to actions, but need not *be* actions or be one-to-one with them[3] - more on this under Act.[4]

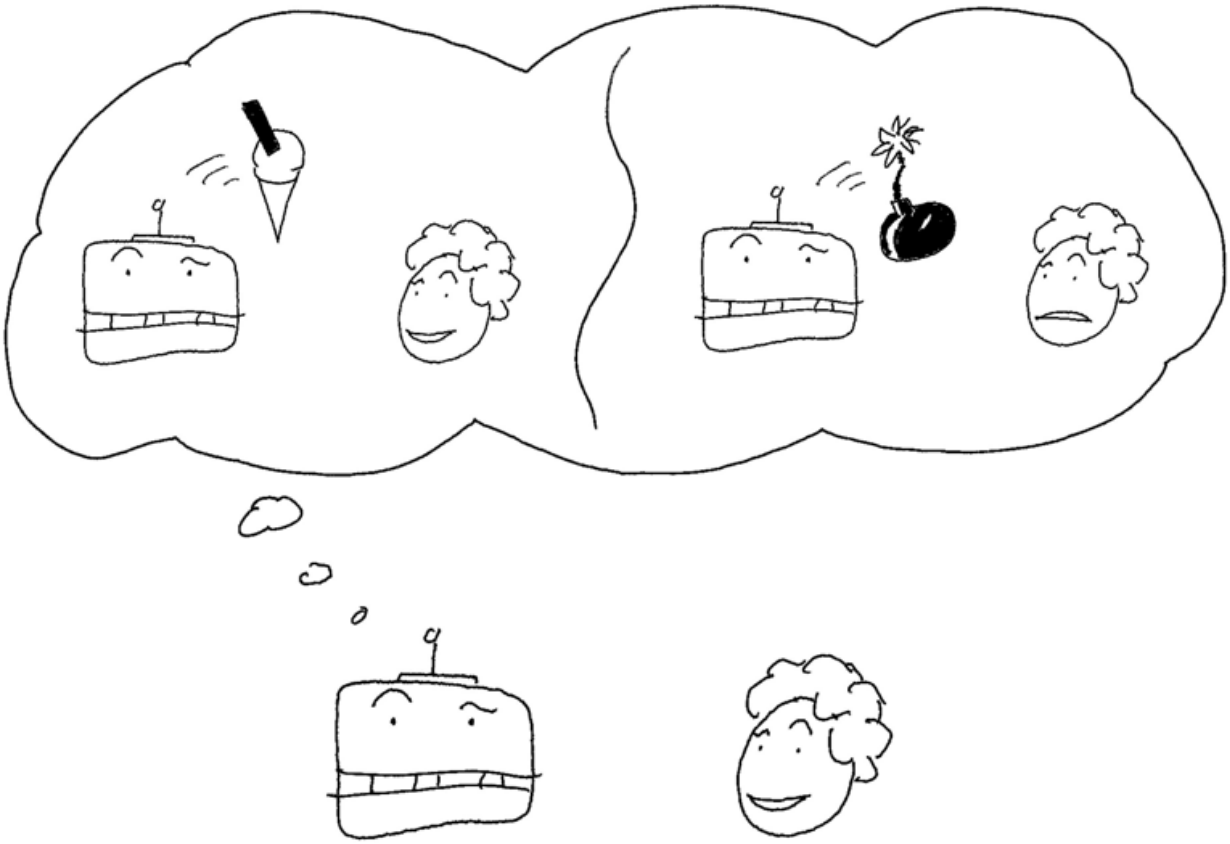
$$\text{Propose} : S \to \Delta\{X\}^{\text{nonempty}}$$

Note one pattern for achieving this is to sample one or more times from

$$\text{ProposeOne} : S \to \Delta X$$

and similarly taking the union of several proposal-collections could be a proposal-collection in some cases.

Some instantiations may have a fixed set of proposals, while others may be flexible.



Propose: robot considers proposals relating to the concepts of delivering ice cream or delivering grenade (NB this image conveys a substantial world model and proposals corresponding closely to plans of action, but this is merely one embodiment of the deliberation framework and not definitive)

Promote

Given a collection of proposals, a deliberative process produces some promotion/demotion weighting V for each proposal. This may also depend on state.[5]

$$\text{Promote} : S \to \{X\} \to \{V\}$$

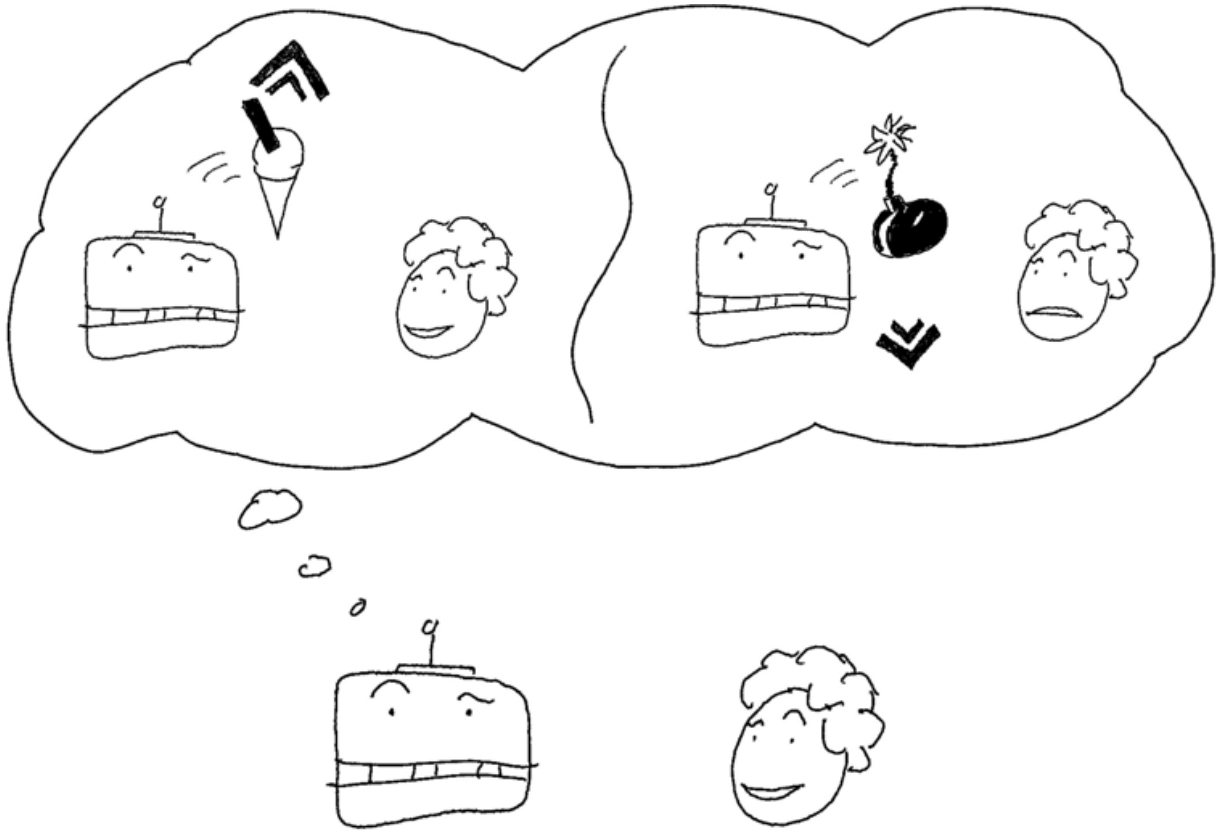
In many instantiations it makes sense to treat the weighting as a simple real scalar $V = R$[6].

Many instantiations may essentially map an evaluation over the proposals

$$\text{Evaluate} : S \to X \to V$$

either serially or in parallel. This suggests 'evaluative' as a useful adjective to describe such processes.

Other instantiations may involve the whole collection of proposals, for example by a comparative sorting-like procedure without any intermediate evaluation.[7]



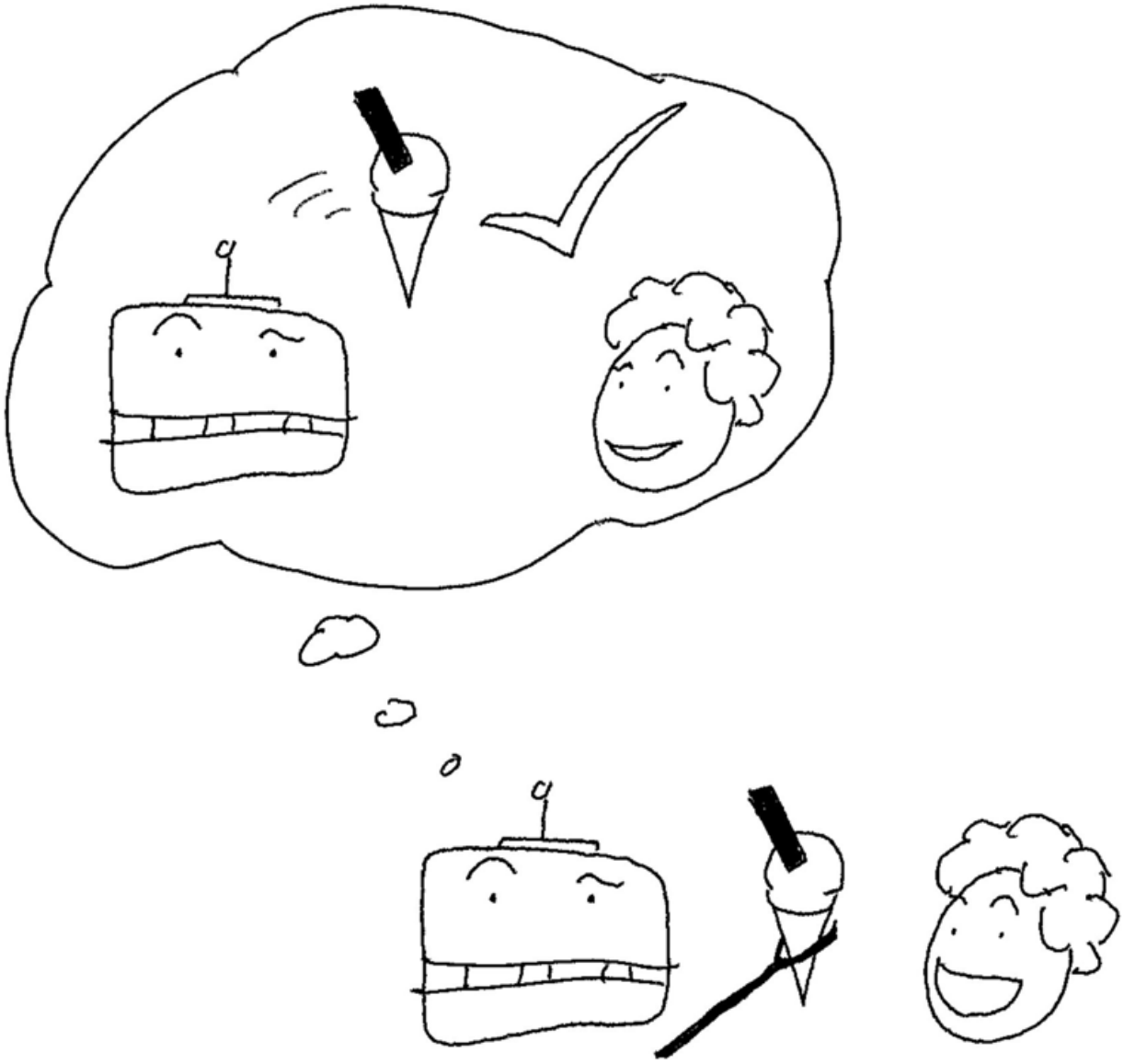
Promote: robot evaluates *proposals of throwing ice cream or grenade respectively, promoting ice cream and demoting grenade*

Act

A decision algorithm eventually acts in its environment. In this analysis, that means taking promotion-weighted proposals and translating them, via some machinery, into activity A.

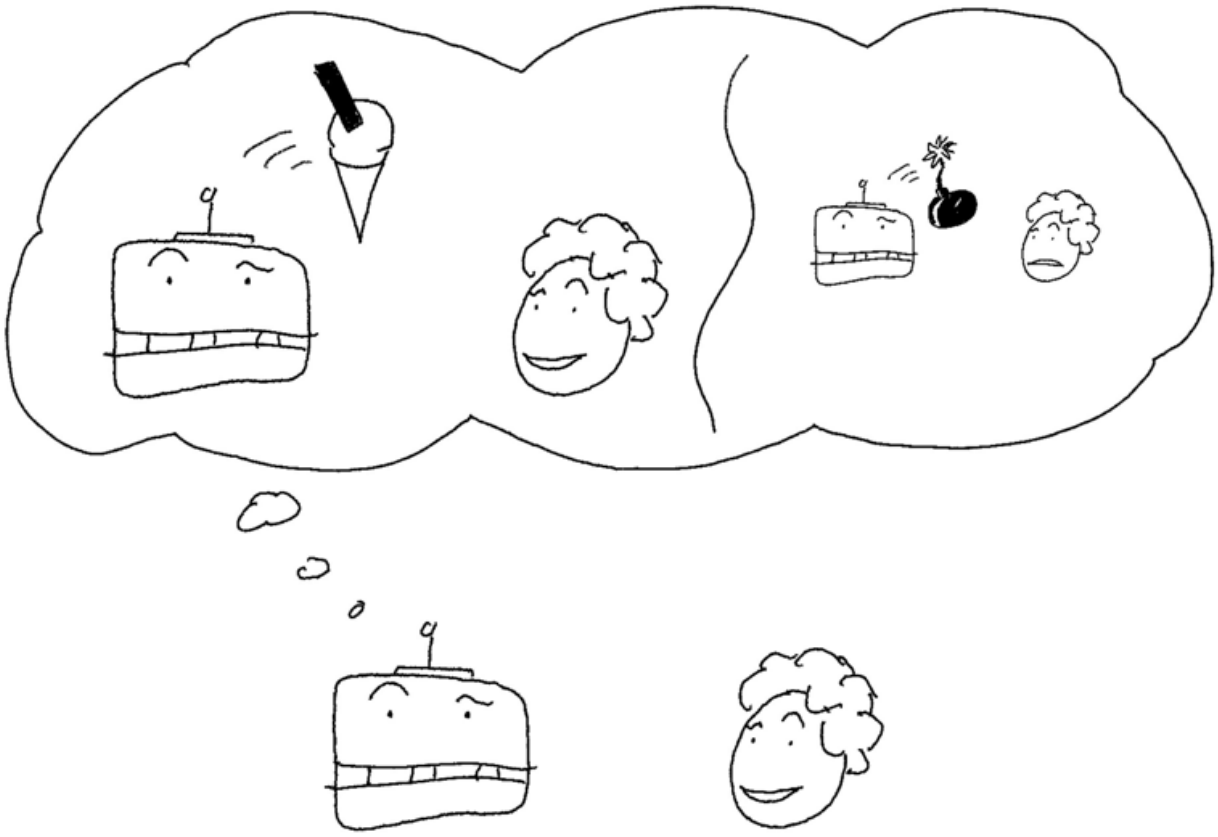
$$\text{Act} : \{X \times V\} \to A$$

In many cases proposals may correspond closely as precursors to particular actions or plans. Promotion may then correspond to preference over plans, in which case action may approximately decompose as $\mathrm{Opt} : \{X \times V\} \to X$ and $\mathrm{Enact} : X \to A$ i.e. a selection of an action, plan, or policy, which is then carried out[8]. Let's tentatively call such systems 'optive'. Enact may involve signalling or otherwise invoking other (reactive or proper) deliberations!



Act-optive: robot's evaluation and promotion leads to opting *to give ice cream, which is a precursor to subsequent motor* enaction, *giving ice cream. Meanwhile the robot may or may not be performing new deliberations.*

More generally, proposals may correspond to configurations, and promotion to weighting-adjustment between those configurations, which play out as activity. We might call such systems '(properly) promotive' in contrast to 'optive'[9].



Act-promotive: robot's promotion and demotion of plans produces mostly 'internal' updates to its plans

Reactions

A reaction is any degenerate deliberation with identically one candidate proposal. In this case the Promote step is trivial or vestigial, and action is effectively fused with the other steps.

In some sense, allowing for degenerate cases of 'deliberation' means that this algorithmic abstraction applies to essentially everything (e.g. a rock is a reactive deliberator which always proposes 'do the thing a rock would').

The important thing is that this abstraction allows us to analyse and contrast *non* reactive 'proper deliberators' with a useful decomposition, as well as to identify and reason about more 'interesting' and less vacuous reactions[10] (while still identifying them as such). e.g. 'Where did this reaction come from?', 'What heuristics/abstractions is it (implicitly) using for proposals?', 'How might this reaction combine with other systems and could proper deliberation emerge?', 'What deliberation(s) does this reaction approximate?'.

Iterated deliberation is generalised control

A decision algorithm eventually produces activity in the environment. Any such activity directly or indirectly alters the relevant algorithmic state of the process, from trivial thermal noise up to and including termination (or modification to some other algorithmic form).

In many cases, the activity typically preserves the essential algorithmic form and it may be appropriate to analyse the algorithm as being recurrent or iterated, perhaps with an approximate Cartesian separation from its environment - the action has some component playing out in the world (which updates and provides new inputs) and some component folding into a privileged part of the world corresponding to the algorithm's 'internal state'. Indeed, in many sophisticated deliberators, the most powerful effect[11] may be on the state or condition (or existence!) of subsequent deliberation(s) and actor-moments.

A system whose activity and situation in the environment *preserve the essential outline* of its deliberation algorithm (perhaps with state updates) de-facto invokes a re-iteration of the same (or related) deliberation procedure. This gives rise to iterated deliberation, which in this analysis is synonymous with control[12].

A basic 'reactive' controller is a system which performs degenerate, reactive deliberation, but which is (essentially) preserved by its actions, giving rise to an iterated reaction, a relatively classic control system.

A (properly) deliberative controller is a deliberative system generating multiple proposals while being (essentially) preserved by its actions, giving rise to a (properly) deliberative control system.

Note that in this sense, what might be considered a single organism generally consists of multiple controllers (reactions or otherwise), and multiple organisms taken together may comprise one controller. Likewise an artificial actor may comprise multiple deliberative and reactive components, and a system of multiple artificial actors may compose to a single deliberative process. The relevant object of analysis is the algorithm.

Recursive deliberation is more general still

Just as the outcome of a deliberative actor-moment may result in zero (terminated) or one (iterated) invocations of relevantly-similar future actor-moments, there is no reason why there should not be a variable number sometimes more than one (homogeneous recursive).

We observe replicative examples throughout nature, as is to be expected in a world where natural selection sometimes works.

As well as invoking multiple copies of 'itself' (or relevantly-similar algorithms) through its actions, as in replication, a deliberator may invoke, create, or otherwise condition other heterogeneous deliberators, as in delegation (heterogeneous recursive). (In fact replication per se is usually not atomic, and goes via other intermediate processes.) Invocation or conditioning of heterogeneous deliberators is what humans and animals do when locomoting, for example[13], and what some robotics applications do when delegating actuation to classical control mechanisms like servomotors. It's also what we see in bureaucracies and colonies of various constructions and scales, from cells to ant nests to democracies.

What factors go into strong deliberation?

Quality of deliberation depends on the fitness[14] of the abstractions and heuristics which the algorithmic components depend on, as well as the amount of compute budget consumed. These are the raw factors of deliberation.

So a deliberator whose Propose generators make better suggestions (relative to some goal or reference) is *ceteris paribus* stronger (for that goal or reference). Likewise a deliberator whose Promote algorithm more closely tracks reality relevant to its goal (is better fit), or a deliberator which can generate and evaluate more proposals with a given compute budget. We might expect good deliberators to 'model' the goal-relevant aspects of their environment as precursors to the state S used in proposal and promotion[15].

No model is correct in all situations but some are more effective than others over a wider range of 'natural' situations. In subsequent posts this framing is used to examine some real examples, locating some relevant ways their abstractions are more or less fit, and some consequences of this. Importantly, quality of Propose generators is emphasised[16].

Quantifying the *magnitude* of the abovementioned *ceteris paribus* deliberation improvements is important, especially in the context of multiple competing or collaborating deliberators, but this is not attempted in detail here.

Convergent instrumental goals and recursive deliberation

In light of the deliberation framing, and focusing on the potential for *iterated or recursive* deliberation, we can restate and perhaps sharpen the convergence of instrumental goals in these terms.

A deliberator is a single actor-moment, finitely capable and neither logically nor empirically omniscient.

For a deliberator oriented to directions or goals which *can not reliably be definitively achieved* in a single round of deliberation, or goals more generally for which *greater success can be expected* by subsequent refinement of action, proposals and actions *which precipitate the existence and empowerment* of similarly-oriented deliberative actor-moments can be expected *ceteris paribus* to achieve greater success.

In this framing, 'self-preservation' and 'goal-content integrity' are special cases of 'pushing the future to contain relevantly-similarly-oriented deliberative actor-moments'[17]. Other instantiations of this include 'replicating'. The prototypical cases from nature are systems which do one or both of *persisting* in relevantly algorithmically similar form (e.g. organisms having nontrivial lifespans or enzymes and catalysts being untransformed), and *replicating or reproducing* relevantly algorithmically similar forms (e.g. genetic elements and simpler autocatalysts directly or indirectly invoking copies to be made, or organisms producing offspring). As noted previously we also see cases

of delegation to heterogeneous deliberators either by conditioning or wholesale creation.

Now, once we recognise a class of deliberator which does well by (at least sometimes) invoking relevantly-similarly-oriented deliberative future actor-moments (whether by persistence, replication, or other delegation), all things equal, the same deliberator would do *better* to invoke *better* such future actor-moments[18]. So how can they be better? In all the same ways as identified already!

Hence if 'self' is the baseline for persistence or replication, actions which induce future deliberations with better expected fitness-to-goal are better, other things equal. This can span the range of

- 'self state updates' tracking locally-relevant information to improve deliberation (memory)
- improving Propose and Promote heuristics more broadly (learning, exploring, play, experimentation)
- more direct 'self' modification including investing more computation into deliberation (cognitive enhancement)
- improving Act efficacy by tuning existing delegation templates or transforming or aligning resources into new and improved (in the preceding ways) copies or delegates (resource acquisition and technological improvement)

Whether to expect a deliberator to discover and/or be capable of acting on any of these kinds of improvements is a separate matter, though evidently any system able to reason similarly to a human is capable of at least in principle apprehending them.

Immediate takeaways

We broke down 'deliberation' into algorithms for 'proposal', 'promotion' (which may be 'evaluative' or 'sortive'), and 'action' (which may be 'optive' or '(properly) promotive'). Combined with the actor-moment framing, we discussed cases of iterated or recursive deliberation. This framing allows us to discuss commonalities and differences between systems and perhaps make more reliable predictions about their counterfactual behaviour and emergence.

Pointing at different parts of this decomposition, we identified various axes along which a deliberator can be improved, additionally rederiving and refining instrumental convergence for iterated/recursive deliberators.

All this is quite abstract and subsequent posts will make it more concrete with examples, starting with relatively simple deliberators, illustrating some more applications of this framing to generating new insights.

Credit to Peter Barnett, Tamera Lanham, Ian McKenzie, John Wentworth, Beth Barnes, Ruby Bloom, Claudia Shi, and Mathieu Putz for useful conversations prompting and refining these ideas. Thanks also to SERI for sponsoring my present collocation with these creative and helpful people!

---

1. So algorithms can not happen 'all at once' and no process can be logically or empirically omniscient, nor can it, in particular, draw 'conclusions', generate abstractions, or otherwise be imbued with information for which there is not (yet) evidence. ↩

2. Alternatively, a lack of proposals could be considered a termination of the process. ↩

3. In principle, the computation performing proposal might have side-effects - *on its own* produce meaningful outcomes or 'actions' (including 'state updates') - but for now this possibility is elided and effects are considered to happen as part of Act. ↩

4. Type notation. Δ can be read 'distribution over', that is, a generative process which can be sampled from, not necessarily anything *representing* a probability distribution per se. {A} means some collection of A. All types are intended to be read as pure, mathematical, or functional - i.e. side-effect free - rather than imperative. ↩

5. NB type notation here uses currying. The resulting V collection is also implicitly dependently-shaped to match the proposals one to one. We might write something like Promote : S → shape → $\{X\}^{shape}$ → $\{V\}^{shape}$ ↩

6. Possibly a central example of *not* mapping to a real scalar might be a committee where not only are 'votes' behaviour-relevant, but the particulars of *who on the committee* voted which way. ↩

7. We might call this 'sortive' in contrast to 'evaluative' but I am not entirely sold on the usefulness of this terminology ↩

8. Consider an example where the promotions are one or other of value estimates, logits, or probabilities over an action space in RL. ↩

9. Natural selection is a canonical properly promotive example.

   It might be most obvious to think about the 'properly promotive' as a relaxation of the 'optive' case. In the optive case, we had some proposals, and depending on the promotion scores for them, we selected exactly one proposal to carry forward. But more generally we might record or apply the promotion decisions in some way, adjusting behaviour, while retaining a representation of a spectrum of proposal-promotion state going forward to subsequent decisions. Going back to natural selection, except in degenerate cases, there is some weighted configuration space of 'proposals' (~genomes or gene-complexes) at any time, which is adjusted by the process. But rarely does it outright opt exclusively for a particular proposal. It seems plausible to me that this proper promotive action is what more sophisticated deliberators do in many cases, also.

   Alternatively, consider that the distinction between 'opting' for one proposal and 'weighting' multiple configurations on promotion is more of degree than kind: 'opting' for exactly one proposal is just a degenerate special case where all other

alternatives are effectively discarded (equivalent to some nominal zero weighting). ↩

10. Like SGD and biochemical catalysis ↩

11. Powerful in the sense of counterfactually most strongly pushing in a goal-directed fashion. It is difficult to bottom-out these concepts! ↩

12. Terminology note: the conceptually important thing is the iterated deliberation.

    Ruby objected to the use of 'control system' because it evokes classical control theory, which is a more constrained notion. I think that's right, but actually evoking classical control theory is part of the reason for choosing 'control': classical control systems are a useful subset of reactive controllers in this analysis (though not the only controllers)! Classical control literature already uses the words 'controller', 'regulator', and 'governor'. Alternative terms if we seek to distinguish might be 'navigator/navigation' or 'director/direction'. We might prefer 'generalised controller/control'. I eschew 'agent' as it carries far too many connotations. I try to use 'controller' to evoke classical and broader notions of sequential control, without necessarily meaning a specific theorisation as a classical open- or closed-loop control system. ↩

13. High level intentions condition lower-level routing and footfall/handhold (etc) placement, which condition microscopic atomic motor actions. ↩

14. Fitness in the original, nonbiological sense of being better-fit i.e. appropriateness to situation or context. For a heuristic this means how much it gets things right, with respect to a particular context and reference. ↩

15. The good regulator and gooder regulator theorems make a Cartesian separation between actor ('regulator') and environment ('system'), and set aside computational and logical constraints. Under these conditions, those theorems tell us that the *best possible* regulator is equivalent to a system which perfectly models the goal-relevant aspects of its environment and acts in accordance with that model.

    For a deliberator to satisfy this, its Propose would need to at least once with certainty generate the best possible proposal (relative to the goal and observed information), and its Promote would need to reliably discern and promote this optimal proposal. In the more general deliberator setting, it is less clear how to trade off computational limitations and there is (as yet) no 'good deliberator' theorem, but this does not prevent us from reasoning about *ceteris paribus* or Pareto improvements to those tradeoffs. ↩

16. I have found many discussions to typically overemphasise what corresponds here to Promote but it should become clear here and in later posts why I think this is missing an important part of the story. ↩

17. It is still useful to distinguish self-preservation from goal-content integrity in cases when there is a good decomposition including a privileged 'goal containing' component or components of the system, reasonably decoupled from capabilities. ↩

18. This might not be as tautological as it seems if there are fundamentally insuperable challenges to robust delegation. ↩

Deliberation Everywhere: Simple Examples

Crossposted from the AI Alignment Forum. May contain more technical jargon than usual.

The analysis and definitions used here are tentative. My familiarity with the concrete systems discussed ranges from rough understanding (markets and parliaments), through abiding amateur interest (biology), to meaningful professional expertise (AI/ML things). The abstractions and terminology have been refined in conversation and private reflection, and the following examples are both generators and products of this conceptual framework.

We previously discussed a conceptual algorithmic breakdown of some aspects of goal-directed behaviour with the intention of inspiring insights and clarifying thought and discussion around these topics.

The examples presented here include some original motivating examples, some used to refine the concepts, and others drawn from the menagerie after the concepts were mostly refined[1]. Each example is subjected to the analysis, in several cases drawing out novel insights as a consequence.

Most of these examples, for all their intricacy in some cases, are relatively 'simple' as deliberators, and I am quite confident in the applicability of the framing. Analysis of more derived and sophisticated deliberative systems is reserved for upcoming posts.

Brief framework summary

We decompose 'deliberation' into 'proposal', 'promotion', and 'action'.

- Propose : $S \to \Delta\{X\}^{\text{nonempty}}$ (generate candidate proposals)
- Promote : $S \to \{X\} \to \{V\}$ (promote and demote proposals according to some criterion)
- Act : $\{X \times V\} \to A$ (take outcome of promotion and demotion to activity in the environment)

We also identify as important whether a deliberator's actions are final, or give rise to relevantly-algorithmically-similar subsequent deliberators (iteration and replication), or create or otherwise condition heterogeneous deliberators (recursive deliberation).

Reaction examples

Chemical systems

Innumerable basic chemical reactions, like oxidation of iron, involve actions which change the composition or configuration of some material(s). For the purposes of this

analysis these are, alone, mostly uninteresting, but serve to illustrate natural systems which do not preserve their essential algorithmic form and thus do not constitute iterated systems.

Some reactions, on the other hand, involve catalysis, wherein some reagents are essentially preserved or reconstituted in the action, producing the seeds of iterated algorithmic reaction, thus basic 'control'.

Despite pushing in a particular 'goal' direction, these systems are reactions rather than proper deliberations, because they occur without computing alternative pathways[2].

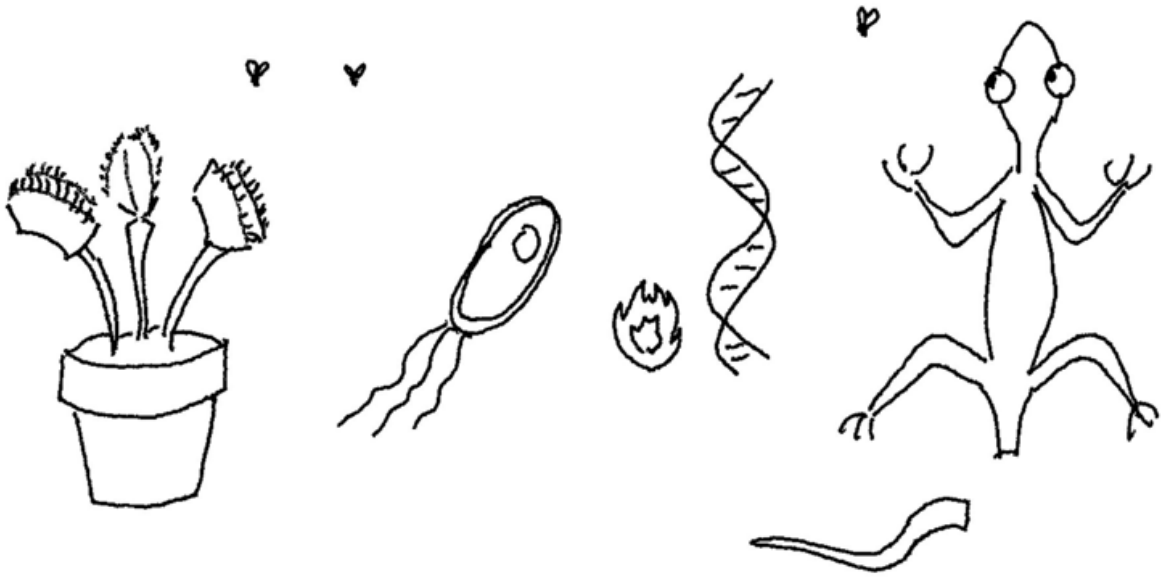
Biological systems

Even very simple organisms can react (that is, act or perform some function in response) to stimuli. No proper deliberation need be involved, no computation instantiated to consider alternatives.

A reaction can take place with or without a brain, or even a nervous system: consider many motions of single-celled organisms[3] or the snapping of a Venus flytrap. The cringe of many animals from intense heat goes via nervous circuitry but takes no deliberation.

Indeed, temperature changes are pervasive in nature, so it is no surprise that we find automatic heat-responsive behaviour at the protein-machinery level in every lineage of cellular life. These latter, along with other protein machinery and the organ-functions of multicellular organisms, demonstrate that, in nature, a single organism will be found to consist of many reactive and deliberative systems.

The class of 'systems undergoing a transformation' in chemical or physical interactions often does *not* preserve the essential algorithmic characteristics of the system, but most salient biological examples are iterated (the act essentially preserves the capacity to further act in an algorithmically similar way). This is not surprising given their origin by natural selection. (Some context-specific counterexamples exist, like some cases of one-shot autotomy to distract predators, though this works precisely because in so-doing, *another* deliberator, the gene complex coding for autotomy, sometimes thereby preserves and propagates its essential form).



Genes and systems-producing-systems

In many biochemical cases, it may be most appropriate to think of active genes and gene-complexes as highly iterated actors, sometimes literally implementing classical control analogues. Their products, the proteins, cells, organs and organisms, can be seen as the mediators of the genes' actions. Some of these mediators are themselves, or give rise to, deliberators or controllers.

The primordial genetic replicators, in the RNA world or even earlier, made little use of mediating systems, but nowadays they predominantly build cellular and multicellular vehicles which are responsible for most of their replicative success. Note that 'creating' is just a special case of 'conditioning', and, once created, cells, tissues, and organs are available to be conditioned as mediators of *other* genes, cells, tissues, and organs' activity.

Of course genes are themselves products and objects of natural selection, so we have systems-conditioning-systems-conditioning-systems.

Artificial reactions

Many artificial systems take reactive actions too. A thermostat adjusts without considering what it might otherwise do, and similar control systems have been in use for centuries at least.

Most contemporary predictive systems, most generative systems, a few game-playing, and many robotics systems produce their outputs reactively - the algorithm is not evaluating multiple proposals, it just proposes some single 'idea' (of varying quality).

There are many contemporary AI systems with rudimentary proper deliberation as well, and some uncertain cases, discussed below.

Many predictive AI systems, e.g. image classifiers, come closest to being 'non iterated' because their results have least bearing on their future invocations (though being more or less apparently useful to human operators does affect this).

Gradient descent as reaction

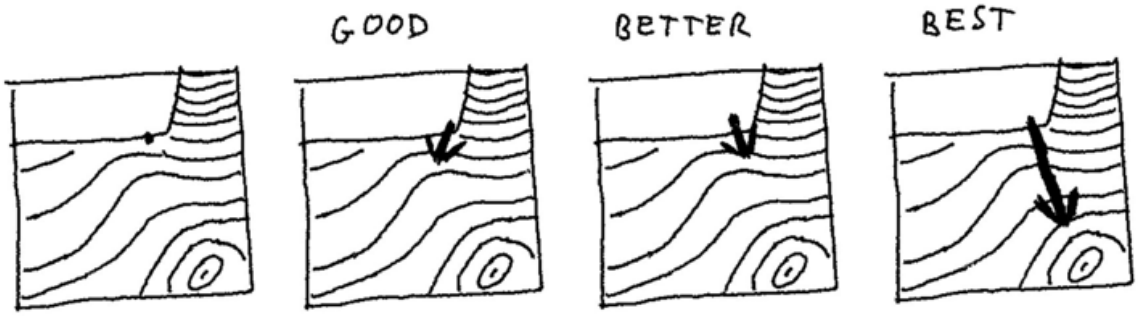
A single step of gradient descent is a reactive motion. A powerful heuristic - compute (an estimate of) the gradient of some target function at a point - produces exactly one proposed update (direction). The update is degenerately promoted and acted upon (applied) in a single fused motion. (In some cases minor proper deliberation may be introduced by evaluating multiple candidate step sizes.) Variations like momentum, and higher-order gradient methods like Newton-Raphson, have the same top-level analysis (though their quality as controllers may differ).

Iterating steps like this results in a controller directed toward ever lower scores.[4]

What makes gradient descent so good?

Despite carrying out no proper deliberation, iterated gradient descent is capable of somewhat-reliable goal-directed control. In fact, under the right conditions, it can produce results just like iterated natural selection which is, in contrast, a canonical *properly deliberative* process (see below). How so?

An essential takeaway from this analysis is that effective evaluation and promotion are *only one part of deliberation strength*. The other part is *coming up with good proposals*. Gradient descent happens to use a strong heuristic[5] for generating good proposals - good enough to entirely compensate for degenerate deliberation, at least when compared to a relevantly-similar natural selection.



Accurate contour rendition of a loss landscape. For navigating to the bottom-right basin from the marked point, a gradient step which is pushed away by the steep slope will eventually subsequently route around, but better steps are possible.

Note that in spite of this clever proposal heuristic, better-still controllers for the same domain are conceivable (though not, perhaps, easy to find or implement), even constricted to reactive deliberation and a fixed learning-rate schedule: *local* best-descent is only the best possible intuition on a globally-linear surface (which is nonexistent in practice). Similarly, appropriately-adaptable learning-rates can proceed more efficiently. This is part of the motivation behind such techniques as momentum, trust region optimisation, and adaptive hyperparameter tuning.

Market arbitrage

At a different level of organisation, arbitrage between sufficiently liquid markets is an example of an emergent reactive system which serves to push toward price convergence, along something akin to a gradient.

The essential form of this tendency is not disturbed by its own action, so it is *iterated*: a controller.

As a robust agent-agnostic process this reaction is constituted by the interactions between *many individual actors*, but is robust to many of the particulars of the behaviour of individuals or subsets of individuals. It serves as an example of a (reactive) controller at the multi-organism level. Even though the individuals enacting the arbitrage may be doing so very properly deliberately, this does not make the arbitrage control process itself any less degenerately reactive.

Deliberation examples

Natural selection, the ur-deliberator

Natural selection with mutation, a weak evaluative promotive proper deliberator, 'proposes' variations on current themes, tries them out by 'evaluating' their success, and moves (on average) to 'promote' fitter combinations[6]. (Natural selection couples its evaluation and promotion to its action, like reactive systems, but other deliberators need not do so.) It is essential to distinguish 'natural selection', the deliberative moment-to-moment or generation-to-generation proposal, evaluation, and promotion/demotion of local variations from '*evolution by* natural selection', the *iterated* deliberative process (controller) which accumulates changes over time.

The implicit abstraction underlying the proposal part of natural selection is basic: sometimes random mutations will be fitter. The evaluation and promotion implicitly depend on the abstraction that if something works, this is evidence that it may work again. These abstractions are weak - the vast majority of lineages of biological organisms are extinct, and most extant organisms carry large amounts of deleterious genes - but we happen to live in a universe where they have been true enough times that life has persisted on Earth for billions of years and adapted to many changing circumstances in that time.

Hyperparameters of natural selection

Certain particulars of the machinery of life on Earth are notable for affecting the 'hyperparameters' of natural selection:

- Separation of concerns into storage (mostly nucleic acids) and function (mostly proteins)
- DNA repair mechanisms, their prevalence and their absence
- Sexual reproduction in many lineages
- Highly-conserved evolutionary development mechanisms

Since none of these can be said to be the primordial state, in some sense it can be said that natural selection has acted on its own hyperparameters. I think it may be appropriate to rather identify 'higher level' natural selections acting on lineages for the

effectiveness of their adaptability, as determined by the hyperparameters of their respective natural selection.

What makes natural selection so bad?

Note that we can turn our question about gradient descent around: what makes natural selection so bad?[7] It uses far more resources than gradient descent but performs comparably.

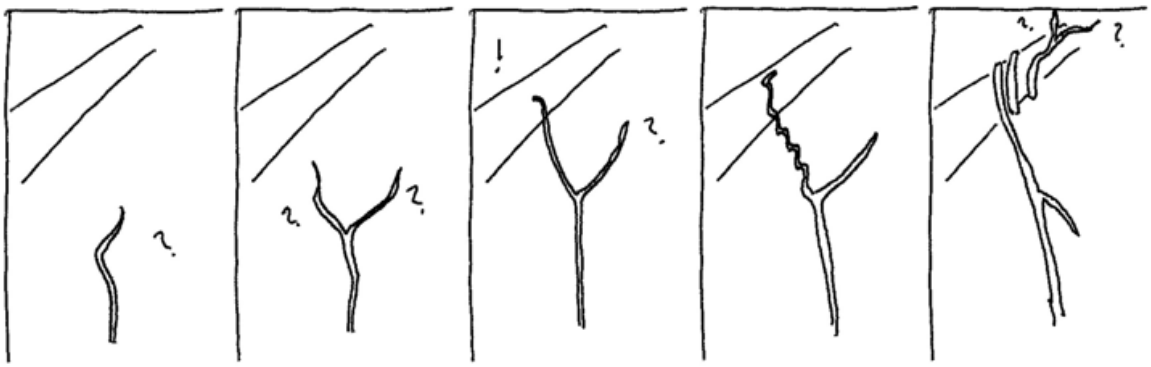
The answer is the mirror image: the proposal-generating and evaluation heuristics of natural selection are about as rudimentary as they could possibly be, so *in spite of* massive parallelism and enormous computational resources, it takes a lot of iteration to get many capability innovations off the ground.

Plants are surprisingly deliberate

Plants as actors are often considered to be lacking deliberation, but in contrast, it is more appropriate to consider many of their behaviours to be weakly deliberative in a fashion comparable to or perhaps surpassing that of natural selection.

For example, climbing plants have numerous flailing, spreading, and otherwise (local) searching routines[8] to generate candidate route-proposals, which, composed with evaluation mechanisms (testing support, light availability, etc) and promotion routines (execute growth, latching, grasping, coiling - when evaluated as favourable) serve as every bit as much a heuristic local deliberative search mechanism as does natural selection.

These procedures are iterated as part of the overall growth control mechanism of the plant, to very successful effect (locating efficient shortcuts to regions of most light and other resources). Other plants, fungi and mostly-sedentary organisms exhibit similar behaviours which should be interpreted in similar algorithmic terms (perhaps especially commonly in roots and hyphae - fungal root-like structures - which are less obviously visible).



Another notable local-search deliberation executed by sedentary organisms concerns their dispersal of offspring and offshoots. Note that for such organisms, conditions as determined by physical location are paramount! It is typical of this lifestyle to invest in many candidate offspring and in a wide variety of dispersal mechanisms. Thus, just as genome is replicated with variation, so is location, and it is therefore part of a kind of

meta-genome eligible to be selected on[9]. As a consequence of this control process, lineages of entirely sedentary organisms can be seen to rapidly reliably proceed toward physical niches to which they are fit. For some species with vegetative propagation it may be more plain that this process is carried out by a 'single actor'[10], but it results in the same kind of 'hill climbing' (or 'shaded fertile valley seeking' as the case may be), because it is essentially the same algorithm.

Why is plant deliberation 'more efficient' than natural selection?

Notably, these control operations, closely algorithmically related to iterated genetic natural selection, nevertheless operate much faster than genetic natural selection (they may take as little as days to manifest 'large' results). Why might this be?

The obvious way to 'go faster' in an iterated deliberation process is to have a shorter iteration cycle, performing more iterations per time. This can account for some of the difference but not all. The other obvious way to 'do better' is to evaluate more proposals per iteration. Notably, natural selection has high parallelism at its disposal and often evaluates *many more* proposals per step than do climbing plants.

It is also important to consider the dimension of the search space, which has important bearing on how tractable it is. The dimension is straightforward: genetic natural selection operates on a high-dimensional space, while plant climbing and locomotion operate on two or three dimensions, so it is simply easier for them to locate good solutions by generating local candidate proposals.

Notice, though, that lichen or mosses climbing trees over time can be seen to instantiate the same search, with similar parameters, in the same dimensions as larger more sophisticated climbing plants, but they achieve this much more slowly (but still fast by comparison to natural selection). So dimension and iteration speed do not appear to tell the whole story.

Two other important properties to pay attention to are suggested by the algorithmic breakdown described previously. First, relating to proposal, the fit of the *sampling heuristics* used in the deliberation. Second, relating to evaluation and promotion, the fidelity of these procedures to tracking the actual target.

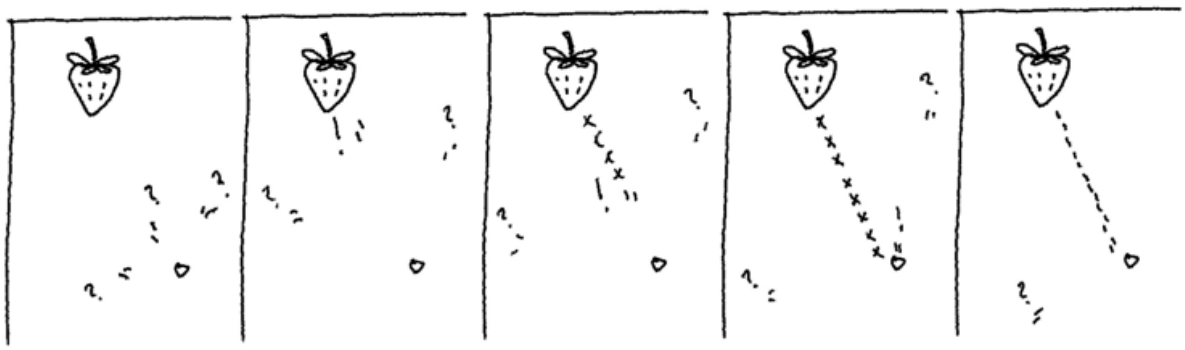
The fit of the sampling heuristics is very important for tractability: genetic natural selection generates its proposals via something similar to symmetrical noise, leading to a lot of dead ends, wasted iterations, and even repeated or approximately-repeated computation. Moss-style migration up trees makes its proposals similarly noisily. In contrast, tendril growth of climbing plants is both decidedly biased (upward) and has an orientation, hence a 'memory', yielding a tendency not to double back on itself, at least locally (unless executing a coil, which is a different part of the algorithm). It is also able to use phototropism to further heuristically bias its search.[11]

These simple but highly-fit heuristic proposal-sampling biases, discovered by natural selection, the slow outer deliberator, may largely account for the evident efficiency advantage of climbing plants over less sophisticated passive climbers like moss and lichen.

Colonial organisms

Viewing insect colonies (for example, ants) as actors, we can identify all of the ingredients of deliberation. The pattern is, like with climbing plants, structurally similar to natural selection, but once again with fitter, more efficient, sampling heuristics.

Consider an ant colony foraging for food. The candidate proposals are the paths taken by individuals or small groups of workers. The evaluations begin with the guesses of those individuals (perhaps augmented by a message-passing algorithm) about the quality and quantity of any foods they come across. The properly promotive outcome involves laying down chemical signals and tactile interactions which encourage attractive or aversive behaviour by other workers (and adjust other behavioural parameters). As an iterated control procedure, this rapidly results in dense lines of workers ferrying the best and most plentiful food back to the nest. Note that no individual ant need be at all deliberative for this to work[12].



This process, again similarly-structured and with similar dimension to the tendril-climbing algorithm of some plants, is able to proceed even faster. From an algorithmic point of view, this can be explained by the even fitter sampling heuristics: for example, individual ants have very strong senses of smell, which allows the foraging party's search to be very heavily biased toward promising locations. This bias is more powerful even than phototropism exhibited by plants, because smell is nonlocal. Another factor is the resource efficiency of the state management: plant climbing algorithms record state and promote candidates by actual organism growth, while ant colonies execute these parts of the algorithm by leaving much less resource-intensive pheromone trails.

Properly deliberative artificial systems

The most obvious properly deliberative artificial systems are training and search algorithms inspired by natural selection, including population based training (PBT) and neuroevolution for neural networks, all of which employ an iterated proper deliberator as a control process to locate and promote complex configurations which are found to score well against some target.

Training and search routines often also employ similar but *non* iterated deliberation, for example static hyperparameter search and random seed search. (Sometimes these themselves end up iterated.)

Other times, the deployed artefact itself is the proper deliberator. Both value-based and policy-based RL controllers over discrete action spaces fall under the properly-deliberative Opt; Enact decomposition in a very basic way. One could make a similar case for a discrete classification model which opts for a single category at the last stage.

Cherry-picking from *reactive* generative models is an example of composing a Promote function (in this case human judgement!) with a reactive system to bootstrap a properly *deliberative* system. There are also automated examples of this type of composition and it is a very general formula for improving decision-making[13].

Hard-coded vs learned deliberation

All of the abovementioned examples of deliberation in artificial systems have the deliberation structure (composing Propose, Promote and Act) coded in by the designers.

- In the case of evolutionary algorithms and PBT, proposals take the form of some carefully chosen local 'mutation' heuristics and promotion is coded around some provided (perhaps approximate) evaluation or fitness function.
- In contemporary non-hierarchical RL with discrete action spaces[14], the proposals consist identically of fixed hard-coded analogues of the atomic action space, the promotion consists of a (usually learned) Evaluate, and Opt; Enact are hand-coded.

- Evaluate consists of either a (state-dependent) mapping from the atomic-action-analogues to estimated utility (expected return, or some similar action-quality estimate, as in Q learning), or to (state-dependent) level-of-intent to take the analogous atomic action (probability or logits, as in policy gradient methods), both of which have $V \subseteq R$.
- Opt translates these evaluations via hard-coded logic into a particular single choice of atomic-action-analogue.
- Enact actuates that analogue in simulation or the real world.

Risks from Learned Optimization as well as prior and subsequent discussions ask questions about when we might expect to find 'optimizers' arising *without* hard-coding them, and about what behaviours and capabilities we should expect from them. Since we have identified *non-artificial* examples of deliberation arising from the refinement or interactions between non-deliberative systems, it seems reasonable to look for analogies in our artificial systems. It remains unclear whether any contemporary artificial systems have learned deliberation.

Further work could consider the question of learned optimization from the deliberation lens, and subsequent posts in this sequence will discuss where deliberation comes from, as well as its manifestation at different levels of sophistication.

Parliaments

Perhaps representing the prototypical 'deliberation', some mechanisms of democratic parliaments illustrate deliberation algorithms at the level of multiple interacting humans.

Consider a vote on some motion, preceded by a debate. This is a proper, quasi-evaluative, optive deliberation.

Even if the motion is indivisible and singular, there are still at least two proposals: enact the motion, or do not enact the motion. In other cases there may be more proposals. These proposals are generated and surfaced by humans within the process by some means or other (which presumably involves its own deliberation!).

The debate, and the sentiments of the individuals participating, are a protracted evaluation and promotion mechanism over the proposals at hand.

The vote precipitates the choice, typically *opting* for some proposal, which is then *enacted* into some derived realisation by one or other organ of the democratic system. (Other possible outcomes to debates are state updates in the form of minds changed or documents written, and re-iterations with new or similar proposals, which is a more properly promotive than optive result.)

Parliaments rarely permanently dissolve themselves (though their actions can lead to their dissolution), so should be seen as iterated deliberators.

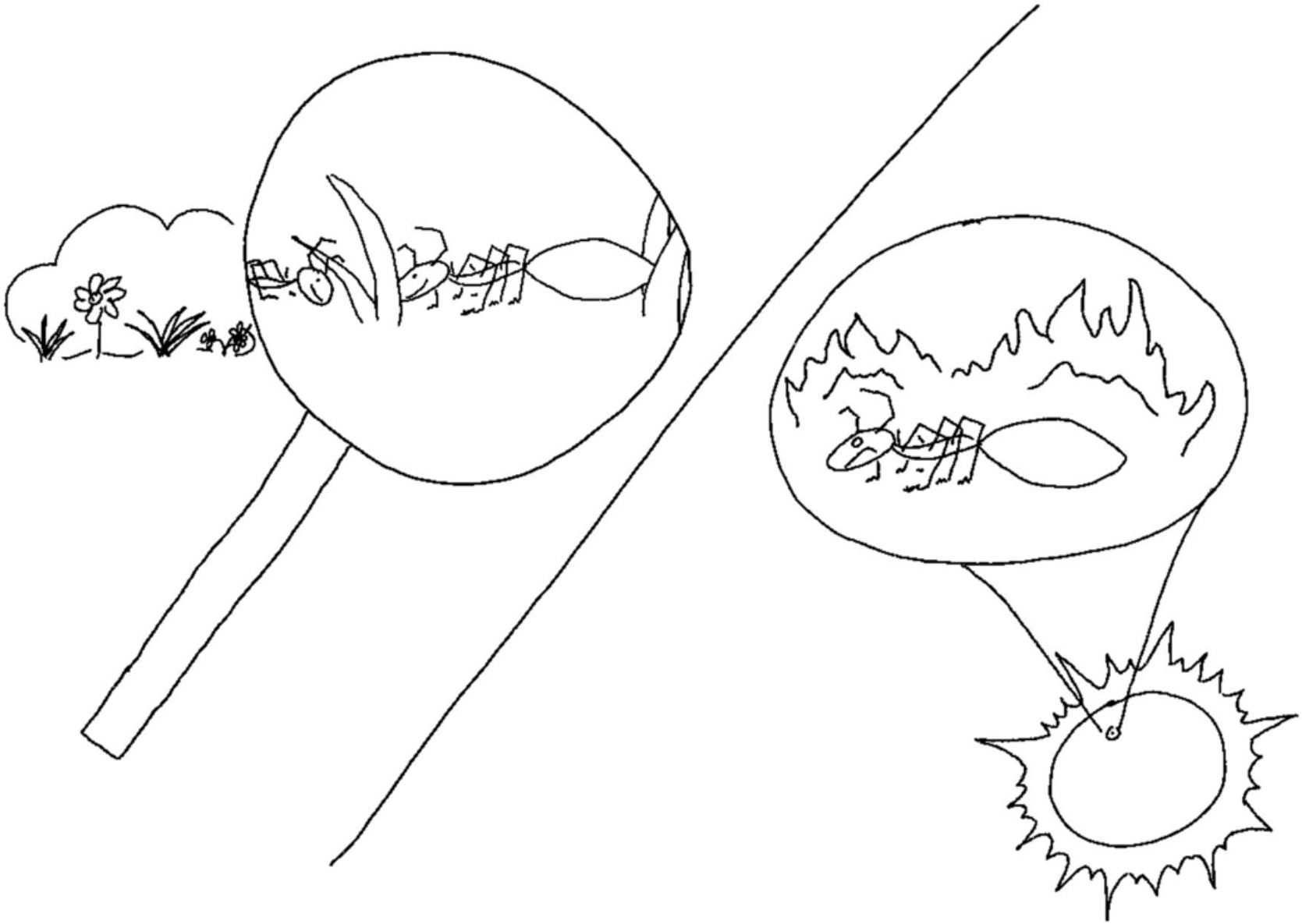
Often the parameters of a particular parliamentary session are set by some invoking organ, body, or individual, making a complex of deliberators which fit together in a larger democratic system.

As in previous cases, the power of a parliament lies partly in producing 'good' proposals, partly in *recognising and promoting* good proposals, and partly in the suitability of its delegate organs to *appropriately enact* its proposals. In the iterated sense, behaviours which precipitate improvements to these capacities are instrumental.

A note on these algorithmic assignments

One may rightly opine that an assignment of a *particular* algorithmic abstraction to a system, hiding the full and intricate detail of every moving piece, must involve some ultimately arbitrary or subjective choices. Perhaps this undermines the preceding discussion.

Indeed, no abstraction can work well in all circumstances, but abstractions are nevertheless inescapably necessary tools to make reasoning tractable in a world with bounded computational resources. If an ant colony were transported to the surface of the sun, there would rapidly be no sense in which it continued to carry out the previously-identified food-searching algorithm. In fact, the abstraction 'ant' would cease to be useful in the same instant. But if the same colony were transported to many locations on the surface of the Earth, or other relevantly-similar hypothetical places, the abstract algorithm identified above would continue to usefully predict outcomes, including counterfactuals ('what if I put some food here?', 'what if I move some ants there?' etc.), at least for some time, depending on the fitness of the colony's algorithm's abstractions and heuristics to the new environment.



'Ant' and 'ant colony' cease to be useful descriptions in the context of the surface of the sun, but remain defensible summary descriptions on many parts of the surface of the Earth. On the left a photograph of a magnification of the author's garden. On the right an artist's impression of the same experiment on the sun. (The author was regrettably unable to attend for the latter experiment.)

Conclusion

We've explored various deliberative systems (both proper and reactive), finding them at many levels of organisation, from cells to democracies. In several cases we uncovered novel insights, such as commonalities between natural selection, climbing plants, and ant colonies, and the deliberation abstraction further enabled us to point to relevant discrepancies determining the 'strength' of some of these systems. Suitability or fitness of Propose appears at least as important as that of Promote for deliberation strength in many cases, exemplified by the comparison of natural selection with gradient descent.

Various deliberative systems comprise interactions between multiple deliberators. These can be seen emerging from the interaction between relatively homogeneous subdeliberators (as in colonies and markets), or in deliberators giving rise to, or conditioning, other heterogeneous delegate deliberators' behaviour (as in cellular and multicellular life and bureaucracies).

Later posts will go into more depth about where deliberations come from, and also examine some deliberators which are more sophisticated, either in how they arrange their delegation or how they perform the key components of deliberation. This includes some artificial systems as well as some of the more cognitively sophisticated and powerful behaviours of animals and humans.

Thanks to those who probed in conversation, and thanks to SERI for sponsoring the research time and making those conversations possible.

---

1. In some sense this resembles a training, validation, and test data split, and in a similar way it gives me some confidence that the concepts here have some explanatory power. The *real test* though is other people attempting to use the ideas or providing criticism. ↩

2. I am not a chemist or physicist but based on my rough understanding of molecular thermodynamics and quantum mechanics, this might not quite be true, and it may be that there are in fact nigh-infinitessimal deliberations happening which 'add up to' what looks like smooth reactions, in a similar way to the right kind of infinitessimal natural selection adding up to gradient descent. ↩

3. Bacterial 'twitching' may be a counterexample which involves basic deliberation; compare below discussion of plant tendrils ↩

4. The iteration here actually comes from an outer weakly properly deliberative evaluative process which always proposes 'step again' and 'stop' and evaluates these against some measure (e.g. validation score, compute budget, some combination ...). Usually 'step again' is promoted by default in a short-circuit loop

to save compute. This outer deliberation is a controller which delegates most of its action to the single-step gradient descent reaction. ↩

5. The heuristic is predicated on the extrinsic (human-derived) abstraction, empirically often applicable, that repeatedly moving, perhaps noisily, straight in the direction of local steepest descent can often effectively find *non-straight* paths to globally-low places. ↩

6. Note that the 'on average' claim takes a stochastic *ex ante* model of fitness as a latent property which is stochastically realised in numbers of progeny. Natural selection can be alternatively understood as just the *ex post* tautology that 'things which in fact propagate in fact propagate'. ↩

7. It feels wrong to say this; I am in fact in awe of natural selection. Perhaps it is a sign of my own bias for inappropriate anthropomorphism that I feel the need to apologise for saying harsh things about an emotionless natural force. ↩

8. Observe two fascinating short time-lapse montages of these literal 'tree searches' from Sir David Attenborough and the BBC at

    - Life: Plants: Climbing plants (https://www.bbc.co.uk/programmes/p005fptt)
    - The Private Life of Plants: Climbing Plants (https://www.bbc.co.uk/programmes/p00lx6cl)

    ↩

9. Note that the Price equation does not care whether a characteristic is described by genes or some other information-carrying attribute ↩

10. The 'humongous fungus', possibly simultaneously the largest-spanning and most massive 'single organism' on Earth, falls in this category ↩

11. It is less clear whether climbing plants with tendril action have improved *evaluation and promotion* heuristics vs moss-style migration, but certainly in both types these heuristics track the target less noisily than does realised-fitness in natural selection. ↩

12. That is not to say that an individual partaking in this collective algorithm *can not* be deliberative; a tribe of individually-deliberative humans could also execute this type of procedure. The deliberative algorithm in this case just happens to be located outside of any single organism. ↩

13. In general, a good enough promoter can expect to get a benefit of a few standard deviations of quality from a proposer this way, though there are rapidly diminishing returns on extra compute spent. ↩

14. For RL or other sequential control with *continuous* action spaces, either the atomic-action-analogues are discretised or sampled from a predefined distribution, in which case the same analysis applies, or the action analogues are sampled directly after a learned policy distribution, in which case we have a learned *reaction*. Hierarchical cases will be discussed in a later post. ↩