

From: Timothy Verstynen timothyv@gmail.com
Subject: Re: eLife rejected; advice sought.
Date: December 21, 2020 at 9:41 AM
To: Erik Peterson erik.exists@gmail.com

TV

Now THIS is what I call good feedback.

On Mon, Dec 21, 2020 at 12:22 PM Erik Peterson <erik.exists@gmail.com> wrote:
Badr, also very helpfully.

(e)

Begin forwarded message:

From: "Albanna, Badr Faisal" <badr.albanna@pitt.edu>
Date: December 21, 2020 at 9:11:09 AM PST
To: Erik Peterson <erik.exists@gmail.com>
Subject: Re: eLife rejected; advice sought.

Hi Erik,

Apologies for the delay. I was swamped with a few other things last week and took a break from my email. Stay tuned for an invitation to join the slack and a link to the git repo.

I'm sorry to hear that this was rejected - as I said when we spoke I found it really interesting and think it has the potential to make a very valuable contribution for the haters as well as the lovers (if only for the haters to defend their position). I know this email is more than you were asking for but I thought it would be more useful than a gut reaction alone.

It might be helpful to restate what I think the big picture idea here is:

* Traditional approaches to exploit/exploit view reward value as the only metric of success. Explore is conceived of as necessary in the face of uncertainty and is defined by stochastic actions that may increase future reward at the cost of current (known) rewards.

* Your paradigm replaces this by two separate values (reward and information) and two corresponding deterministic searches (the latter is referred to as "curiosity") for each type of value. A win-stay-lose-switch strategy is used to decide whether to switch from pursuing reward value to information value (and vice versa). (A boredom cutoff is used to decide when information is insufficient to prize curiosity)

* Information value is measured as the amount of change induced by an observation on the memory state (defined as a geometric norm).

* An optimal deterministic search is possible.

* When implemented this algorithm outperforms other common strategies in accumulating reward value on many different standard tasks (even though it is not set up to optimize reward value alone).

I can't speculate where the theorist got lost, but I can tell you the places that required a re-read for me. Some of them did leave a queasiness in my stomach that I didn't quite shake and I imagine for a reviewer with less time to look at this they may have been dealbreakers. I admit that some of this confusion might be a result of my relative lack of depth in the reinforcement learning literature but then again maybe the reviewer had the same problem.

1. The notion of information "value" snuck up on me in line 59 and it wasn't immediately obvious that it was a new type of value unrelated to reared value. In my mind the clearest presentation consistent with the overall goal here is to state early that you are introducing a new type of value independent of reward which is tied to how much an observation changes your internal "memory" or "representation" of all observations through the current one. In that spirit it may be useful to state your own definition of traditional reward value before hand.

3. I really had a hard time understanding how all of this was possible without specifying the structure of f . For example definitions 1 and 2 on lines 99 and 100 seemed to really be about how you should specify f rather than a definition of E . Maybe I'm thinking about this wrong but in the trivial case where f maps everything to the same point, no curiosity is possible. I just couldn't get past how one should choose your f - presumably there is an optimal choice here which represents the "right stuff" but then it seems like this is a way of sneaking reward value into your algorithm. Maybe I'm overthinking this but it made it hard to understand how you made your results - did it work so well because of your choice in f ?

3. Relatedly, I found it odd that "memory" requires forgetting but not recall. To me this is another place where perhaps the two types of value are getting tied (memories must be stored in a way that facilitates action for reared seeking). Not sure if you are already thinking of it or implementing it in this way but it would go a long way toward helping me understand you are bringing this down from the level of the abstract to implement it.

4. By banishing uncertainty, it seemed like your algorithm had to know the impact of every possible observation on memory.

7. By banishing uncertainty, it seemed like your algorithm had to know the impact of every possible observation on memory before taking an action. If I'm getting this wrong I would definitely like to know. Seems implausible but I may be thinking about this wrong.

Two more thoughts regarding the comments you shared

Biological relevance: Honestly I think this has huge implication for how we view reward signals in the brain. Seems like you model that these two measures of value have to be comparable in some way (perhaps they use the same neurotransmitter / circuits etc) and there must be two operate mechanisms/circuits/readouts for each of the value systems. This is a massive prediction in my mind and it is worth looking into if your picture can make an argument about some of the debates regarding dopamine, acetylcholine, etc. function. To me even a speculative paragraph or two here would really go a long way towards justifying this being presented to eLife's audience.

Different structures for memory & observations: One last thought related to the role of "f" above has to do with the structure of the space. Obviously a bounded vector space or embedded manifold is only one possible structure for observations & memories and one of the advantages of your approach (to me) is that any algorithm for seeking value on a structure works just as well for reward as it does for information. That said, I'm again left wondering how much these choices affect your results but that's obviously another story.

Hope some of this was helpful!

B

—

[Badr F. Albanna](#), PhD (*he/him/his*)

Faculty

Department of Neuroscience

University of Pittsburgh

Crawford Hall Rm. 449

<http://albannalab.science>

[@AlbannaLab](#)

On Dec 15, 2020, at 5:24 PM, Erik Peterson <erik.exists@gmail.com> wrote:

I could use some advice from someone who can follow the math, I hope (see below), and who has no reason to be nice.

You, Badr, in other words.

The main editor did a mini-review of other editors, and sent it to an outside theory person for their comments. In the end, the editors felt it was not biological enough or different enough from other work.

Which, fine.

It is the comments that were passed on from the outside reviewer which worry me. And it is here I hope you'll confirm or deny, without reservation.

> The new theoretician who looked at it had some strong reservations too, mostly that (despite their very extensive and broad math/computational background), they felt "lost", not because of the math of any given equation but they felt the overall presentation was "incomprehensible", and they could not corroborate the conclusion that a new optimal solution had been found. They also noted that alternating between seeking information and seeking reward and trading them off is confusing since these have different units. They felt that "if I'm lost, I'm guessing the average elife reader will also be lost."

Does this to you ring true?

The editor also said,

> I do have to agree that the presentation could be improved, but moreover that the scope of the paper is not well suited for eLife without more specific connections to biology

Does this to you ring true?

(PS - the currency thing I can address. It is not a new question.)

I don't need another deep read from you. I'm asking off the cuff, right now, what do you think?

--

(e)

--

Timothy Verstappen, PhD (*he/him*)

timothy verstyken Ph.D. (ne/ning)

Associate Professor

Psychology, Carnegie Mellon Neuroscience Institute, & Biomedical Engineering (by courtesy)

Co-Director of the CMU-Pitt BRIDGE Center

Carnegie Mellon University

Pittsburgh, PA 15213

www.cognitiveaxon.com