Review of Taking fear back into the Marginal Value Theorem

I very much enjoyed this manuscript. It attempts to unify two previously separate subfields, a goal I consider valuable and important. The paper opens with excellent motivation -- even though I am not a specialist in either MVT or Brown's GUD frameworks the authors convinced me that unifying the two is work worth doing. Much of the mathematics was well explained, the results appear sound, and I appreciated the intuition they provided behind many of them.

Because I am not a specialist I will focus my review on large picture issues and narrative flow. I did not check many of the technical mathematical derivations (my stochastic theory is embarrassingly bad).

Major Issues

I see only a few "major" issues. I suspect none are fatal, but only parts in need of clarification. However, some of these parts are for core ideas of the paper and so I am calling them "major."

MVT Graphical Analysis and concavity of the gain function

The paper often mentions that one can apply the same graphical analysis as in MVT. My understanding is that these graphical methods only work or give unique solutions if certain technical conditions hold (ie. the gain function being concave or some similar shape that you can find the tangent line). It is not obvious to me that the transformed gain functions meet these requirements. I would like to see more discussion on this and proofs that the graphical analysis can always be applied.

This graphical argument appears similar to that of the Levin's fitness function/set method, which is also used in the Smith-Fretwell model. It is an adjacent field, but still might be worth citing.

The paper left me unsure of how important is it that the gain function for each patch is concave? It seems to me essential to the argument as a few derivations rely on Jensen's inequality. I would like to see expanded discussion around this, whether non-concave change the results, and whether such functions exist in nature.

Additive partitioning

(Line 296) Is additive partitioning is provably impossible to achieve? Or were the authors unable to find a way? I'm reading it as the former but if so I would like to see the proof. I imagine the proof is fairly trivial, but as this is an important claim I think its worth writing down in the supplementary materials.

The whole distinction between Charnovian and Brownian formulations and what exactly is meant by the partitioning are also worth discussing further. I would like the authors to expand on the significance that only one case can be written in Brownian form. What does this mean for this larger program of unifying Charnovian and Brownian views? What does this mean biologically?

Mathematical Terminology and Notation

I found the terminology and notation to be fine overall, but I ran into certain problem areas which I think could be improved for clarity.

I felt the term "transformation" along with the tilde notation and $\tilde{F}$ are overloaded with different meanings.

- In equation 2 we have the "transformation of the gain" and in equation 4 we have the "transformed gain." Are these different names meaningful? I thought not on first read-through as they both use the same notation. But then it becomes confusing on whether 4 is building on 2 or is a completely separate case.
- My current understanding is that $\tilde{F}$ in equation 2 is the "real life" gain function when we add in risk, escape, and other details. That is, it is a better approximation of reality than the original gain function F. I feel this use of the term "transformed" conflicts with "transformed time" and the transformed gain of equation 4, both which feel more of a technical mathematical manipulation.

In equation 5 we have this dot notation defining the function. I have not seen this notation before and I'm confused about how the function is defined. My current reading is that it is typoed and means $G_j(\tilde{t}) \equiv \tilde{F}_j(\tilde{t}^{-1})$. Furthermore, it seems that F is called the gain function, $\tilde{F}$ is the transformed gain, and G is back to being called the gain function. This is confusing.

Sometimes the patch subscripts are dropped. I am confused on what this indicates. β **is a particular offender.** I could not find where it was defined yet it is of importance later one, especially when defining optimal boldness.

The paper uses risk-MVT and rMVT interchangeably. Sometimes both versions are used in the same paragraph (eg Table 1 caption) making it appear they refer to different things. I would choose one and toss the other. (My vote is to keep risk-MVT and it is clearer and still fairly concise.)

Overall I would recommend the authors revisit some of this confusing terminology, and provide a table of notation. With the six different cases there are enough moving parts that a table will go a long ways towards readability.

Optimal Boldness

I found the distinction between boldness and risk confusing, and whether these are per-patch or global definitions. It seems that boldness to me is βt while risk is β (which I remark again is undefined!).

But in that case isn't optimal boldness is just the optimal foraging time rescaled? Once you know the optimal time t^* you gain no new information by considering optimal boldness. It feels more like a definition, one which I do not understand the significance of.

To me optimal boldness has a connotation of deciding how risky of a patch you are willing to visit. Ie. something the forager controls separate from time spent. By the importance given to optimal boldness in the title and abstract I expected it would be a deeper and more discussed aspect. Instead it seems quickly gone over and insignificant.

I would like to either see boldness have a deeper treatment, or for the title and abstract to change to deemphasize this. Personally, I think some title that makes it clear the authors are unifying Brown's and Charnov's frameworks would be better, as that seems to be the deep significant result and the soul of the paper.

Small things

Here are some minor points. Some of these are important but easy fixes. Others are small nitpicks. Many are likely due to my unfamiliarity with the subfield, so I defer to the authors whether these are worth changing.

- (160) I'm likely being dense, but I do not see how this is a linear rescaling. Did the authors mean it is approximately linear for high values of β_j? I would appreciate clarification.
- I'm confused about the distinction between transformed time and effective time.
- Would be nice to define what a micromort is within the text. I do not think this is familiar to many readers, but seems easy to define.
- "It is quite universally true in biology that organisms do not perceive time uniformly, but rather 391 in a relative way, depending on internal state and external factors."
    - Is it? It sounds believable but its is still a strong claim and I would like more citations.
- Equation 1
    - The MVT is specifically saying that the solutions of equation 1 maximize individual fitness, correct? I found this a little unclear. Perhaps putting line 84 right before the equation can help.
    - (75) It would be nice to mention what the n in E_n stands for to help mnemonicly. Later on I forgot and kept thinking it was an index term like $j = 1, . . , n$.
- Figure 2
    - The quotation marks in (b) both face the same direction. (Actually, this is true for all figures.)
    - I would appreciate slightly more detail for the caption of (b) since it could serve as a nice reference for what each of the six elementary combinations mean. Alternatively, we could have a table reference for the combinations.
        - I really liked how the authors considered these six combinations as they seemed very natural, but I'd like the reference as I found some of the names difficult to remember.
- Paragraph 131
    - Is there any reason to call these states (i) and (ii) rather than the foraging and interrupted states? The latter is easier to remember and read and I see no advantage on the former.
    - I am curious to why the return to foraging rate γ is not based on patch type? It seems a reasonable simplification but I would like just a line or two of justification.
    - Does δ scale with how long the interruption lasts? Or is it a one off payment for being interrupted. I'm reading the former in the equations but the text is slightly ambiguous.
- Figure 3 + 4
    - (Figure 3) The formatting on the plots seems to have exploded.
    - (Figure 4) What do the blue lines mean? Is this meant to show the one-to-one mapping?
    - (Both figures) What do the grey lines mean?
- Figure 5
    - Is the "all or nothing" blue line covering the orange? Or does the effective timescale not exist here?
- (Table 1) Could the caption clarify what is meant by rMVT domain?
- Figure 5 typo "I each case..." I found three or four other typos two I forgot the location of. All minor, but maybe worth a copy edit pass.
- I found the introduction well written, clear, and concise. The conclusion I felt was less so. It's not bad, but the quality of the introduction spoiled me and I suggest another revision pass on the conclusion.

I listed many issues in the name of constructive critique, but I want to emphasize again that I greatly enjoyed this manuscript and the ideas in it. It is an excellent paper with a significant contribution. I hope others like it as much as I did.