


Reviews For Paper**Paper ID** 4356**Title** Learning to select computations**Masked Reviewer ID:** Assigned_Reviewer_1**Review:**

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	The paper introduces a new approach to metareasoning in which a linear combination of approximations is used to construct a metalevel MDP which is solved to give the metareasoning policy. This is demonstrated to give good performance on 3 synthetic examples.
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Moderately significant
[Novelty] Are the problems or approaches novel?	Novel
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Sufficient
[Clarity] Is the paper well-organized and clearly written?	Satisfactory
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	<p>This paper addresses an interesting problem and demonstrates progress. While the results are entirely experimental and on very simple synthetic problems, the approach is novel and interesting.</p> <p>The biggest weakness of the paper is that the conceptual part (pages 1-3) is too terse and painful to follow. I would gladly have sacrificed the first example domain from the evaluation for more gradual, leisurely, and detailed explanation.</p> <p>The second main weakness is that there are no interesting theorems and the experimental evaluation is on entirely synthetic simple problems. The third problem domain is a multi-step decision process - the term "planning" is used but the actual problem doesn't reflect any planning problem I'm familiar with (the agent can learn the cost of any edge in a tree through deliberation, without learning anything about its</p>

	<p>predecessors). But on balance I think the conceptual contribution is sufficient to carry the paper over the acceptance threshold.</p> <p>minor things:</p> <p>Q^*_{meta} is not defined (eq 4)</p> <p>B_{t+1} (eq 6) is not defined (as opposed to b_{t+1})</p> <p>"that confers the highest improvement" - really? why limited to myopic greedy?</p> <p>emphasize that eq 8 has the E inside the U^\wedge</p> <p>the 3-dimensional optimization doesn't seem "simple"</p> <p>is it easy to be slightly more specific about the BO procedure? anything special about it or could something else easily have been used to the same effect?</p> <p>less than 10 -> fewer than 10</p> <p>not quite sure why you are calling the object-level problem an MDP if it is deterministic.</p> <p>the closing discussion section comes across as pompous and vague. For example, "the foundational" -> "a foundational". I would also suggest omitting speculation.</p> <p>The claim about "wider range" seems false - blinkered performed pretty similarly, didn't it?</p> <p>Puterman's book was originally published in 1994 (giving the date of the reprint is perhaps misleading)</p>
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	<p>To a cynic, calling this RL is just giving a fancy name for the fact that you tuned the weights until it worked well. Is it really a surprise that you can surpass greedy with such a scheme? It's nice to see that it's possible (and I would argue that it's worth publishing), but I'm curious if you see this work as having any lasting contribution? </p>
[OVERALL SCORE]	Accept
[CONFIDENCE]	Reviewer is knowledgeable in the area

Masked Reviewer ID: Assigned_Reviewer_2

Review:


Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	<p>The paper proposes an approach, based on Reinforcement Learning, to learning to select computations with the view to optimising the benefits regarding decisions based on the computations, and at the same time minimising the cost of the computation. In experiments, the proposed approach compares favourably with existing methods.</p>
[Relevance] Is this paper relevant to an AI audience?	<p>Likely to be of interest to a large proportion of the community</p>

[Significance] Are the results significant?	Significant
[Novelty] Are the problems or approaches novel?	Novel
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Sufficient
[Clarity] Is the paper well-organized and clearly written?	Good
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	The paper makes a contribution towards solving an important and difficult problem. The work is solid and according to the presented experimental evidence the results seem to be significant. The presentation is rather theoretical and so are the problems studied in the experiments. Maybe the paper could be made easier to read by including real-world examples which would also make practical motivation for this work more concrete.
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	The example from chess about the effectiveness of human computation is good, but it is very far from the scope of the actual technique proposed in this paper. On the other hand, the examples of problems at the end of the paper, to which the proposed techniques might possibly be applicable, are stated very generally. Can you think of more concrete examples to motivate the practical relevance of your work?
[OVERALL SCORE]	Accept
[CONFIDENCE]	Reviewer is knowledgeable but out of the area

Masked Reviewer ID: Assigned_Reviewer_3


Review:

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	This paper presents a method for learning how to meta-reason, i.e., to select which computation to perform in order to gain new information, or when to stop computing because the additional rewards do not justify the computational costs. The method is based on reinforcement learning of the weights of a linear interpolation between different approximations of the value of a computation and its cost. The approach is evaluated experimentally and shows promising performance.
[Relevance] Is this paper relevant to an AI audience?	Likely to be of interest to a large proportion of the community
[Significance] Are the results significant?	Significant

[Novelty] Are the problems or approaches novel?	Novel
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Very convincing
[Clarity] Is the paper well-organized and clearly written?	Good
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	<p>The paper is quite well written (although some of the notation is confusing to readers who are not immediately familiar with the previous work). It clearly addresses an important problem that is highly relevant for many sub-areas of AI. The novelty in this paper is to have a bounded-optimal approach, and to introduce a concrete set of features that can be learned for the approach to be tractable in practice.</p> <p>From what I can tell, the paper is technically sound (I am not an expert in this area), and it contains a thorough experimental evaluation of the proposed method that clearly demonstrates its merits.</p> <p>In terms of presentation, especially for readers who are not deeply familiar with the related work, the notation is sometimes confusing. For example: equation (4) refers to Q^*_{meta}, which is not explained; equation 6 and other equations later refer to B which has not been introduced (I am wondering whether this is supposed to be b); the definition of VOC_1 (and VOI_1) refers to Russel and Wefald 1991a, however the notation in that paper differs quite a bit from this one, and I could not find an explanation (in terms of equations) of the meta-greedy policy or the myopic VOC. Improving these aspects could make the paper more accessible to the general AAAI audience.</p> <p>Typos etc: Page 1, "that can achieved by" Eq (5), $VOC_1(b,c)$ should be $VOC_1(c,b)$ Eq (6), it's unclear what c_t refers to, and should B_{t+1} be b_{t+1}? Page 1, "Russel & Wefald, 1991" is that 1991a or 1991b (same in footnote 1) Page 6, "tthrough"</p>
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	Please clarify the notation. 
[OVERALL SCORE]	Accept (Top 50% accepted papers (est.))
[CONFIDENCE]	Reviewer is knowledgeable but out of the area

Masked Reviewer ID: Assigned_Reviewer_4

Review:

Question	
[Summary] Please	This paper proposes a  algorithm for metareasoning. The algorithm uses

summarize the main claims/contributions of the paper in your own words.	<p>various estimates of value of information as features in a model to deduce the value of computation. The authors then show across three domains that their algorithm produces achieves higher utility than state-of-the-art algorithms.</p> <p>There are a number of issues specific to the field of metareasoning which must be addressed, but are not. The most serious issue is the time that it takes to run the proposed metareasoning algorithm. Any metareasoning algorithm must be extremely quick, because otherwise, in the time it takes to metareason, you could just solve the original problem. The authors claim in a very hand-wavy fashion that their algorithm is efficient. However, there are no details (what algorithms are actually being used, how they are being used, and what the computational complexity of those algorithms is). While the authors show that their algorithm works better than the meta-greedy heuristic, their evaluation does not consider the time it takes to run the algorithm. It seems from the evaluation that their algorithm requires a substantial amount of offline training (which the meta-greedy heuristic does not). Could the underlying problems be solved in the time it takes to do the offline training by some basic planning algorithms?</p> <p>Connected to the above point and also important, the authors are conflating "metareasoning" with receiving observations about the world. For example, in their first example domain, their metareasoning algorithm needs to decide whether to receive another observation from the world or to take an action based on the observations that have been seen. This is not really metareasoning. Metareasoning would reason about whether to run the algorithm that would determine whether or not to take another observation from the world or take an action based on those observations. Of course metareasoning boils down to a planning problem, but the framing is such that it must reason about computation, and reason about it quickly.</p> <p>More detailed comments below.</p>
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Moderately significant
[Novelty] Are the problems or approaches novel?	Novel
[Soundness] Is the paper technically sound?	Has minor errors
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Not convincing
[Clarity] Is the paper well-organized and clearly written?	Good
[Detailed Comments] Please elaborate on your assessments and	<p>"The computations C have no external effects." This is a strong assumption. The world does not stop because you are thinking. Doing nothing at all may be better than doing something. For example, deliberation may prevent you from buying a stock that just dropped.</p>

provide constructive feedback.	<p>Typo in equation 5 - b and c are reversed.</p> <p>I don't understand equation 8. b is the current belief. Shouldn't the first term of the value of perfect information be given by the optimal policy with the true parameters θ?</p> <p>There is a lot of handwaving with respect to how things can be computed. "VOI_1, VPI, VPI_A can all be computed efficiently or efficiently approximated by Monte-Carlo integration" How, and how efficiently? "We use Bayesian optimization to optimize the weights in a sample efficient manner" Again, please explain how and how fast?</p> <p>Experiments in general: Why is there an arbitrary horizon? Why shouldn't you allow the agent to continue to reason if the value of computation never drops below 0? What if you just ran some monte-carlo based pomdp algorithms? How well would these do in the same time it takes to train the algorithm you propose?</p> <p>It's not accurate to call the third experiment "metareasoning about planning". Metareasoning about planning involves reasoning about the planning process (the solving of the planning problem). In this experiment, that would mean the process of figuring which path to take down the tree <u>after</u> the rewards are learned. Given all the rewards of the tree, the agent must still plan to figure out which path to take, and there is no reasoning about this.</p> <p>It is unclear to me why the proposed algorithm is a reinforcement learning algorithm. What is the exploration/exploitation tradeoff in the algorithm?</p> <p>Some additional relevant papers: Horvitz and Breese, Ideal partition of resources for metareasoning Horvitz, Cooper and Heckerman Reflection and action under scarce resources: Theoretical principles and empirical study.</p>
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	see above
[OVERALL SCORE]	Reject
[CONFIDENCE]	Reviewer is knowledgeable in the area