**Semantic Visual Navigation by Watching YouTube Videos.** We thank the reviewers for their thoughtful comments. We are glad that the reviewers share our excitement about the paper, and found it "valuable for the AI community" (R2), "highly relevant to NeurIPS" (R4), "brings a new perspective" (R4), and one that "could open up new research avenues" (R2). Reviewers also found our methodology "novel and interesting…showing great promise and outperforming a plethora of baselines" (R2), "empirical evaluation well-conducted and convincing" (R1), with "credible ablation studies and analysis" (R2). Reviewers also found the paper to be well-written (R1, R2, R4). We address comments from R1, R2, and R4, and will appropriately incorporate them in final version. Unfortunately, R3's review misrepresents our work in multiple ways (see below). We point out these factual errors, and urge the AC to view R3's review in this light.

**Response to R1.** **New Goal Objects**: Our model can't adapt to new objects at test time, but given a detector for the new class, we can re-train without extra annotations or environment interactions. **Outdoors**: Videos in $\mathcal{V}_{\text{yt}}$ have outdoor shots (patio, yards, which we currently filter out), thus our models, when retrained, could work well in those outdoor contexts, but not as well elsewhere. We can't evaluate as Gibson doesn't have outdoor environments. **Use of alternate algorithms for short-term navigation** is possible, as our policy is modular. Though, as noted, it is orthogonal. **Fixed Size Heap?**: Yes! Current episodes only popped $\leq 25$ items. Larger envs / farther goals would need a larger heap.

**Response to R2.** **Comparison to ObjectGoal SOTA**: "Active Neural SLAM [11]" doesn't tackle ObjectGoal (but only free-space exploration & PointGoal). Our Strong Supervision Value Function (L291) can be thought of as an ObjectGoal extension of the more recent CVPR 2020 "Neural Topological SLAM [12]". When trained on 85 environments from $\mathcal{E}_{\text{video}}$ with strong supervision, it achieves an OS-SPL of 0.528 (Tab. 1). When combined with $D_{\text{coco}}$ this goes up to 0.547. In preliminary experiments done for rebuttal, jointly training this strong supervision value function with our method on $\mathcal{V}_{\text{yt}}$ in a multi-task manner achieves an OS-SPL of 0.563 ($p$-value of 0.098 over the 0.547). Thus, learning from videos leads to improvements on top of recent methods even when they use strong supervision in 85 environments.

**Response to R4.** **Component design choices, lack of ablations**: We already provide ablations in Sec. 4.2 (and Tab. S6 in Supp) where, as suggested, we swap in ground-truth versions for each component (in $\mathcal{V}_{\text{syn}}$, as $\mathcal{V}_{\text{yt}}$ is unlabeled). Using ground truth versions only lead to a 1%–3% improvement in final performance. This makes us confident in our design choices. We will augment these existing experiments with the rationale for our design choices. Inverse dynamic model was trained on *held-out* 15 envs and worked well (acc. 95%), more details on L223, L195, Tab. S1 in supp.
**Pre-training in RL**: Great point. We already tried initializing RL policies with ImageNet trained models (L269). That said, this still ignores $\mathcal{V}_{\text{yt}}$ data, and test-time access to detector $D_{\text{coco}}$ used by our method. We present two experiments that control for these. First, we give the RL models access to $\mathcal{V}_{\text{yt}}$, by initializing with the model trained using BC in L280 (which in-turn used ImageNet initialization). This doesn't help much (OS-SPL of $0.24 \pm 0.02$). Second, we test our model without $D_{\text{coco}}$. This achieves an OS-SPL of 0.44, *vs.* 0.29 for the best of RL models. Thus, our approach isn't just "piggy-backing off of the pretraining." We will include these additional experiments in the final version.
**Improvement over Detection Seeker**: **1.** Under the tighter *paired student t-test*, SPL (standard metric for ObjectGoal [4]) of our method ($\mathcal{V}_{\text{yt}}$ version) is better than Detection Seeker with $p$-values of $0.0006$ and $0.068$ in Oracle and Policy Stop settings respectively. Tab. 1 reported 90% confidence intervals (*i.e.* a looser, unpaired test) so as to report all methods together. **2.** Improvements over Detection Seeker are more evident in hard episodes (where agents starts far from target object), as seen in SPL breakdown across episode hardness (Sect. S2.2 & Tab. S2 in supp). It is difficult to improve upon Detection Seeker in easy cases when the object is likely already in sight from around the start point.
**Other**: Trends are similar at success threshold 0.5m (OS-SPL: Ours 0.34 / Det. Seeker 0.31). Note, in policy stop setting behavior is *completely* autonomous. Visualizations in Fig. 3,S3,S4,S7 show what our model learned. We will add generalized value function references, correct the potentially misleading phrasing, & further discuss broader impact.

**Response to R3.** "**…central contribution (as noted in L119) is a value function …**": L119 explicitly notes that our novelty is in *the use of videos* for learning value functions. This differentiates it from [1.r], other prior works [28,42,65].
"**…not novel compared to [39]. In particular, …topological map is very similar to [39, 49, 6].**": This is wrong, [39, 6] do not even use topological maps!! [49] does, but for a different task (going to an image goal in a pre-explored environment), and without the high-level semantic value function as we do. [39] is related for a different reason. L98 explicitly describes this relationship, which R1 notes as being "fairly discussed."
"**…no prior IfO approach included as a baseline in the experimental studies …**": Our Behavior Cloning on Pseudo Labeled Videos (L280) is precisely the BCO(0) algorithm from prior IfO work [57]. We achieve relative improvement of $33\% - 250\%$ SPL over this baseline. We will note this relationship to [57]. References [2.r–7.r] fall in exactly the same category as IfO references [5,18,57,58] discussed on L105: tackling the same task in the same environment depicted in the demonstration *vs.* our work that solves novel tasks in novel envs. Furthermore, [2.r,5.r,6.r] obtain policies through RL on reward functions learned from task demonstrations. Thus, their performance is upper-bounded by that of using dense ground truth rewards (already in Tab. 1, OS-SPL of 0.29, *vs.* 0.50 for Ours). We will cite [1.r–7.r].
"**…marginally better than behavior cloning …Detection Seeker performs competitively …**": $33\% - 250\%$ relative improvement in SPL over behavior cloning (0.24 *vs.* 0.50, 0.06 *vs.* 0.21, 0.36 *vs.* 0.48, 0.10 *vs.* 0.21, Tab 1.) isn't marginal!! Detection Seeker performs competitively, but our gains over it are significant (see L33 above).