

1 **Overview.** We thank reviewers for their constructive feedback and clarified the paper based on their comments. This
2 paper introduces language goal imagination as a new exploration mechanism, and presents a systematic study of its
3 impact on exploration and generalization. We were pleased that reviewers acknowledged the novelty and interest of this
4 approach. **R2** valued "the *extremely* thorough empirical evaluations", **R5** the experimental extensibility, and **R1** and
5 **R4** the strong experimental evidence supporting our claims on exploration and generalization.

6 **Motivation for a descriptive setup and "real-world problems" R1.** Our setup is motivated by a developmental
7 approach which investigates how child development can inspire the building of autonomous agents. Social interactions
8 play a key role for children: they mostly follow their own goals and get descriptive feedback from adults (see L114,
9 **R1**). Our descriptive setup mimics these findings, and thus differs from the classical instruction-following setup: agents
10 set their own goals, imagine goals and receive language descriptions (*descriptive feedback*) instead of success/failure
11 signal (*instruction feedback*). Descriptive feedback can have real-world applications. It helps sample-efficiency and
12 facilitates human labeling with *a posteriori* counterfactual feedback (irrespective of the selected goal) and is especially
13 useful when providing instructions is not straightforward (e.g. possible effects are unknown). We are currently working
14 on an industrial application leveraging such descriptive setups for digital assistants interacting with real users in smart
15 homes. It is also more user-friendly to comment the actions of an autonomous system than giving instruction that may
16 fail. We will mention this applied work in the discussion.

17 **Why do we need the Playground environment? R1** Existing benchmarks (like BabyAI) were missing two features:

- 18 1. To our knowledge, existing benchmarks use instruction feedback whereas we want to study descriptive feedback.
- 19 2. In Playground, agents can ground the meaning of object categories and the meaning of category-dependent dynamics
20 (animals grow with food or water, plants grow with water only). We can thus study new types of generalization (see
21 Supp. Sec. 2). We believe that the full power of goal imagination occurs in environment with rich combinatorial
22 dynamics involving a wide variety of objects.

23 Playground is a tool that we hope will enable the community to further study under-explored descriptive setups with
24 rich combinatorial dynamics, as well as goal imagination (as noted by **R2** and **R5**). The environment can easily be
25 extended by adding objects, attributes, category- or object-type-dependent dynamics (**R2**) and will be open-sourced.

26 **How to scale Playground to more goals? R2, R4 and R5** were concerned about the size of the set of feasible goals
27 (256) (out of $9 \cdot 1e6$ possible sentences, **R2**). However, they include different types of interactions (navigation, grasping,
28 growing), whereas traditional approaches usually only consider navigation to a target object (e.g. Doom [12], DMLab
29 [32]). The possible set of goals would easily grow by introducing synonyms. Using pre-trained language models would
30 be useful there (Lynch et al., 2020). Having a small set of goals allowed us to easily control the complexity of our
31 environment. We believe the study of new mechanisms (here goal imagination) benefits from controlled environments,
32 rigorous methodology and statistical testing (supported by **R2**).

33 **How can I adapt (R2) or extend (R5) IMAGINE to other RL approaches?** Our descriptive setup was studied in
34 isolation from instruction-feedback, but the two can be combined. Two features of IMAGINE could be added to
35 traditional RL approaches (**R2, R5**). First, descriptive feedback can be used in any setup that assumes a descriptive
36 function of the trajectory (either hard-coded [12] or learned [17]). Second, agents can imagine goals as long as they
37 learn a reward function that can generalize their meaning from known goals ([5], as noted by **R1**). Descriptive feedback
38 would facilitate human labeling (e.g. Lynch et al., 2020) and improve sample efficiency (via hindsight learning). Goal
39 imagination would help with generalization and exploration as demonstrated by our experiments.

40 **R5** wondered whether the grammar construction process would be needed in a new setting. In new environments,
41 reusing learned patterns is likely to boost further exploration: this suggests interesting future work. In more complex
42 language settings, [1] showed how similar mechanisms yield fruitful results for data augmentation on NLP datasets.
43 More realistic language could also be mapped to simpler language to facilitate goal imagination (Andreas et al., 2020).

44 **Other comments.** In our approach, sentences are goals. Testing goals are built from the same atomic words (null
45 atomic divergence), but have maximum compound divergence [41]: they are *out-of-distribution* w.r.t. the distribution
46 of goal sentences from the training set (L168, **R4**). Generalization is evaluated by the average success rate (SR) on
47 goals from the testing set (**R5**). We provide a potential explanation to the surprisingly high I2C of the low coverage
48 condition in Supp. 1-251 (**R2**). We did not pair evaluations and thus did not use paired t-tests (**R2**). Substitute goals
49 g_s can belong to \mathcal{G}^{train} or \mathcal{G}^{im} (**R2**). Vygotsky showed that infants use more egocentric speech for planning when
50 tasks become harder (Chap 2 - *Thought and Language*, **R2**). We agree the hierarchical algorithm presented in [36] is
51 complementary to ours (**R4**). IMAGINE boosts low-level exploration by imagining novel goals and could be extended to
52 a hierarchical setting (example of downstream task, **R1**). Related work on instruction following in more complicated
53 visual environments (Das et.al 2017, Nguyen et. al 2018 and Shridhar et.al 2020) will be discussed (**R2**).