Lessons from a "failed" replication:
The importance of taking action in exploration

Hashem Sadeghiyeh[1,*], Siyu Wang[1,*], and Robert C. Wilson[1,2]

[1]Department of Psychology, University of Arizona
[2]Cognitive Science Program, University of Arizona
[*]equal contribution

November 9, 2018

Abstract

In recent years, psychology has been rocked by the findings from a series of large scale studies in which classic, and even textbook, findings have failed to replicate. This 'replication crisis,' has lead to damaged careers and thrown who subfields of study into doubt. In part because the stakes are so high, a number of authors have pushed back on the notion that psychology is in fact in crisis, instead suggesting that failed replications may be due to unknown 'moderator' variables that vary between labs. Despite the intuitive appeal of this argument, relatively few examples exist where a failed replication has been 'explained away' in this manner. Here we present such a case in which a replication failed and through further analysis and a follow-up experiment we discovered a previously unknown moderator variable with high impact on the phenomenon we have been studying. More specifically, the phenomenon is the 'exploration': exploring unknown options in search of information, and the moderator variable is 'acting' to obtain information versus passively receiving it. Moreover, we show that by switching between active and passive versions of the task during a single session, we can turn on and turn off the effect within subject. While this single finding hardly settles the replication debate, it illustrates the point that, as with any experiment, the results of a "failed" replication must be interpreted with care.

Keywords— exploration, exploitation, explore-exploit dilemma, failed replication, active, passive, learning; attention, description-experience gap, action readiness

Introduction

The replication crisis has hit psychology hard. In one famous attempt, Nosek et al. (2015) were unable to replicate findings from over 60% of studies published in high impact journals in 2008. Similarly, Camerer et al. (2018) failed to replicate findings from 7 of 21 psychology papers published in Science and Nature between 2011 and 2015. More generally, this lack of reproducibility has led to damaged careers (e.g. Sayre & Riegelman, 2018), concerns about whole subfields of research (Earp & Trafimow, 2015; Pashler, Coburn, & Harris, 2012) and has hurt the credibility of psychology in eyes of the public and experts (Ferguson, 2015).

In part because the stakes of the debate are so high, the notion that there actually is a replication crisis has received considerable push back (Pashler & Harris, 2012; Maxwell, Lau, & Howard, 2015; Gilbert, King, Pettigrew, & Wilson, 2016). While these authors raise multiple concerns, the issue we focus on here is the criticism that the whole process of attempting a replication may end up rewarding bad science. It seems easy to get an experiment to fail, by simply failing to conduct the experiment in the way in which it was designed (Stroebe & Strack, 2014; See (*xkcd: Negative Results*, n.d.) for a comic demonstration of this concern). That is, even when a replication is attempted in good faith, a failure to replicate could be down to minor differences in implementation that dramatically change the results. Despite the appeal of this argument in theory, in practice there are relatively few examples in the literature in which a failed replication can be 'explained away' by a previously unknown, moderator variable (Simons, 2014). In this paper we present one such example where a seemingly minor change in the task led to a profound change in the results.

The particular behavior we focus on is a type of decision problem known as the explore-exploit dilemma. In explore-exploit decisions, we must choose between exploring actions whose reward is unknown and exploiting actions whose reward is well defined. For example, when dining at a restaurant, do you choose a familiar item with guaranteed satisfaction (exploit) or do you pick a new item on the menu which might be more delicious, but you don't know until you try it once (explore)?

Recently we presented evidence that humans use at least two distinct strategies to solve the explore-exploit dilemma: directed exploration and random exploration (Wilson, Geana, White, Ludvig, & Cohen, 2014). In directed exploration, exploratory choices are driven by information whereby people prefer more informative options over less informative options with the same (immediate) expected value. Conversely, in random exploration, exploratory choices are driven by random decision noise which causes lower value (and often more informative) options to be chosen by chance.

Our evidence for making this claim came from behavior on a newly designed gambling task, known as the 'Horizon Task'. In this task participants choose between two simple gambling machines (one-armed bandits), receiving a reward from the bandit they have chosen. The rewards vary from trial to trial around a fixed average, which is different for each bandit, and with a fixed variance, which is the same for both bandits. In every game, one of the bandits has a higher average payoff than the other, but the subject can only learn which option has the highest average, and hence is best to exploit, by exploring first.

The Horizon Task has two key manipulations. The first is the time horizon, the number of trials participants will make in the future. The idea behind the horizon manipulation is that it should affect how likely people are to explore: with long horizons favoring exploration, as people seek out information that can be useful in the future, and short horizons favoring exploitation, as people maximize their immediate reward. The second manipulation was the use of forced-choice trials, or example plays, in which participants were forced to play the bandits in a particular order for the first four trials of each game (Figure 1). This manipulation allowed us to carefully control the information participants had before making their first free choice. The key finding in this work was that both information seeking and behavioral variability increased with horizon, which we interpreted as evidence for people using both directed and random exploration in the task.

In the original version of the Horizon Task, which we refer to as hereafter as the 'active' version of the task, the forced-choice trials were actually played. That is, during the forced-chioce trials, participants were instructed which side to choose, pressed a button to choose that side and then saw the outcome from the forced choice. Shortly after the senior author started his new lab at the University of Arizona, we attempted to replicate the experiment in this new population. In order to reduce the duration of the experiment, we removed the constraint that the forced trials actually be played and instead designed a 'passive' version of the task in which forced-trial information was presented all at once to participants. We assumed that behavior in this passive version of the task would be no different from the active version of the task because participants would have exactly the same information when they made their first free choice. The only difference in this new design would be in how they got the information from the forced trials: one at a time in response to a button press, in the active version, or all at once, in the passive

replication. To our horror, when we analyzed data we found that the replication had failed and that the horizon effects on both directed and random exploration had almost completely disappeared.

At the same time as we were failing to replicate the Horizon Task results in our own lab, our collaborators were successfully replicating and extending our original findings in their own labs at Princeton (Krueger, Wilson, & Cohen, 2017), Harvard (Somerville et al., 2017) and the Nencki institute in Poland (Zajkowski, Kossut, & Wilson, 2017). In light of these robust replications, we realized that there must be something in the new task design that significantly changed behavior and our suspicion fell on the passive nature of the replication task. To test this suspicion, we thus sought to replicate the behavior in the active version of our task at the University of Arizona and, if this active replication was successful, to more fully explore the effects of passive forced trials within subject.

Results

Here we report the results from our initial failed replication in the passive version of the Horizon Task (Experiment 1) and our follow up study that included both active and passive versions of the task within subject (Experiment 2).

Both experiments are based on the Horizon Task (Figure 1, see Methods for full details). In brief, in this task participants play a series of Games lasting either 5 or 10 trials in which they must choose between two one-armed bandits. When played, each bandit pays out a reward sampled from a Gaussian distribution. The mean of the Gaussian is different for each bandit, but the variance is the same for both.

To control the amount of information people have before making a free choice between the options, the first 4 trials of each game are instructed trials in which outcomes from specific bandits are shown to the subject. In the active version of the task from our original paper (Wilson et al., 2014), participants actually play the instructed trials, pressing the left and right buttons on the keyboard to reveal the outcome of the instructed plays one by one (Figure 1A). In the passive version of the task, participants are simply presented with the outcomes of the four instructed trials all at once (Figure 1B).

The instructed trials are used to set up one of two information conditions (Figure 1C): an unequal (or [1 3]) information condition in which one bandit is played once and the other three times, and an equal (or [2 2]) information condition in which both bandits are played twice. These information conditions allow us to easily quantify information seeking and behavioral variability in a model-free manner. Specifically, information seeking is quantified as the probability of choosing the more informative option in the [1 3] condition, p(high info), and behavioral variability is computed as choosing the low mean reward option in the [2 2] condition, p(low mean) (Figure 1C). Crucially, by measuring how p(high info) and p(low mean) change between the long and short games (i.e. long and short horizon conditions), we can quantify directed exploration as

$$\text{directed} = p(\text{high info,horizon 6}) - p(\text{high info,horizon 1}) \qquad (1)$$

and random exploration as

$$\text{random} = p(\text{low mean,horizon 6}) - p(\text{low mean,horizon 1}) \qquad (2)$$

Using these measures, we find that both information seeking and behavioral variability increase with horizon consistent with the idea that people use both directed and random exploration in this task. In our initial replication attempt we tried to replicate these findings in the passive version of the task.

Experiment 1: Failure to replicate in the passive version of the task

We analyzed the Horizon Task data from a total of 371 participants (292 of which were in the active condition and 79 in the passive condition) in a between-subjects design.
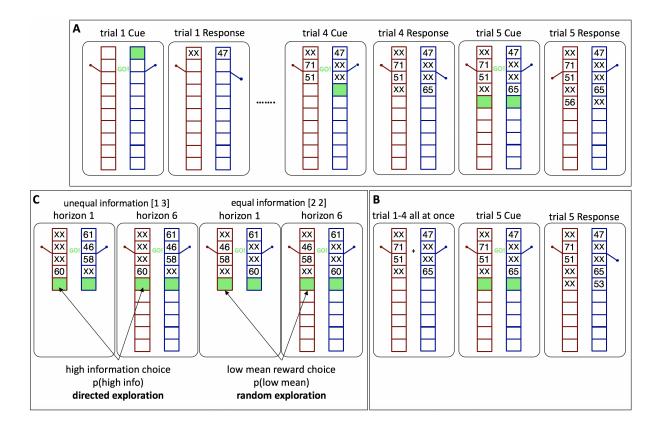
Figure 1: Active and passive versions of the Horizon Task. In the active version of the Horizon Task (A), participants actually play the four forced trials at the start of each game. In contrast, in the passive version of the Horizon Task (B), the outcomes of the four forced trials are presented all at once at the start of the game. In both active and passive versions of the task (C), the four forced trials set up one of two information conditions ([1 3] and equal [2 2] information) and two horizon conditions (1 vs 6) before participants make their first free choice.

Overall Performance

Overall performance in the two conditions is shown in Figure 2. In Figure 2A, we plot the subjects' accuracy as the fraction of trials in which the participants chose the objectively more rewarding option (i.e. the option whose generative mean reward was higher). In general, accuracy was above chance in both active and passive conditions (t-tests vs 0.5, passive, horizon 1: $t(78) = 11.71$; $p < .001$; passive horizon 6 $t(78) = 9.60$, $p < .001$; active horizon 1: $t(291) = 24.78$, $p < .001$; active horizon 6: (291) $= 19.93$, $p < .001$). In the horizon 1 games there was no difference in accuracy between the active and passive conditions (two-sided t-test, $t(369) = 2.25$; $p = .02$). In the horizon 6 games, an ANOVA with trial number and active-passive condition as factors showed a main effect of trial on the accuracy, as accuracy increased over the course of each game ($F(5, 1845) = 62.15$; $p < .001$), but no main effect for the active-passive condition ($F(1, 1845) = .21$; $p = .65$), and no significant interaction between the active-passive condition and trial ($F(5, 1845) = 1.49$; $p = .19$).

In Figure 2B, we show the reaction time over the course of each game. In horizon 1 games, reaction times were slower in the active than the passive condition ($t(369) = -2.42$; $p = .02$). Similarly, in horizon 6 games reaction times were slower in the active than the passive condition as well ($F(1, 1845) = 7.66$; $p = .006$). Consistent with our earlier work (Wilson et al., 2014), we see that participants responded more quickly as the game progressed ($F(5, 1845) = 271.62$; $p < .001$). In addition, this change over the course of the game was more gradual in the active than the passive condition (as indicated by an interaction in
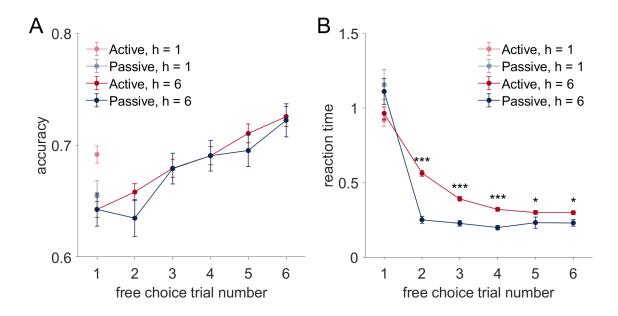
Figure 2: Basic performance in the between-subjects design (Experiment 1). Accuracy (A), defined as the fraction of trials in which the participants chose the objectively more rewarding option and reaction time (B), the interval between onset of a trial and subject's response, through the free-choice trials in horizon-6 games.

the ANOVA: $F(1, 1845) = 16.31$; $p < .001$).

Directed and random exploration are absent in the passive condition

To allow fair comparison between the horizon conditions, our main analysis is focused on the first free-choice trial (the fifth trial) in each game. We used the model-free measures $p(\text{high info})$ and $p(\text{low mean})$ as indicators of directed and random exploration (Figure 1C). In both cases, we used a mixed-design repeated-measures ANOVA with condition (active vs passive) as a between-subjects factor and horizon (1 vs 6) as a within-subject factor.

For $p(\text{high info})$, we found significant main effects for active vs passive ($F(1, 369) = 58.03$; $p < .001$) and horizon ($F(1, 369) = 19.47$; $p < .001$) and a significant interaction between horizon and active/passive ($F(1, 369) = 3.94$; $p = .048$) (Figure 2A). Post-hoc t-tests showed that the main effect of active-passive condition was due to a reduction in $p(\text{high info})$ in the passive case in both horizon conditions (for horizon 1: $t(369) = 5.59$; $p < .001$; for horizon 6: $t(369) = 7.26$; $p < .001$). Meanwhile, the interaction effect was caused by a reduction in the horizon-based change in $p(\text{high info})$ in the active condition ($t(291) = 6.67$, $p < .001$), while this horizon-based change was absent in the passive condition ($t(78) = 1.64$; $p = .10$).

For $p(\text{low mean})$, we found a significant main effect for horizon ($F(1, 369) = 33.15$; $p < .001$) but not for the active-passive factor ($F(1, 369) = 1.31$; $p = .25$). Also there was a significant interaction between horizon and the active-passive condition ($F(1, 369) = 14.68$; $p < .001$). Post-hoc t-tests showed that this interaction was driven by a selective decrease in $p(\text{low mean})$ in horizon 1 in the active condition ($t(369) = 2.26$; $p = .02$). In addition, as with $p(\text{high info})$, the horizon-based change in $p(\text{low mean})$ that was present in the active condition ($t(291) = 10.12$; $p < .001$) was absent in the passive condition ($t(78) = 1.21$; $p = .228$).

As noted earlier, the increase in $p(\text{high info})$ in horizon 6 games compared to horizon 1 ones, is used as an indicator of directed exploration and the increase in $p(\text{low mean})$ in horizon 6 games comparing
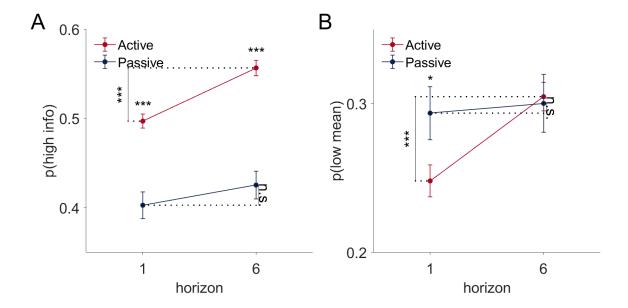
Figure 3: Behavior on the first free-choice trial in the between subjects design (Experiment 1). Both directed (A) and random (B) exploration increase between horizon 1 and horizon 6 in the active condition, but not the passive condition. In addition, participants in the passive condition are more ambiguity averse in both horizon conditions (A) and more variable in their behavior in horizon 1 (B). $*$ denotes $p < 0.05$, $***$ denotes $p < 0.001$.

to horizon 1 games, is an indicator of random exploration. Taken together these results suggest that both directed and random exploration are absent in the passive condition (Figure 3), and the horizon-dependent behavior does not replicate in the passive condition.

Behavior on later trials converges between active and passive conditions

Figure 4 shows $p(\text{high info})$ and $p(\text{low mean})$ through the later trials in horizon 6 games. For $p(\text{high info})$ (Figure 4A), it is significantly larger in the active condition than in the passive one at trials 1 and 2. However, when it goes further, there are no significant differences in $p(\text{high info})$ between active and passive conditions and the non-significant difference becomes less and less as the games go on until it is literally the same in two conditions on trial 6. For $p(\text{low mean})$ (Figure 4B), the difference between active and passive conditions never reached significance level in all trials

Experiment 2: Active-passive manipulation within subject

Experiment 2 was designed specifically to investigate the effect of the active versus passive factor within-subject. Overall, a total of 68 subjects participated in the experiment.
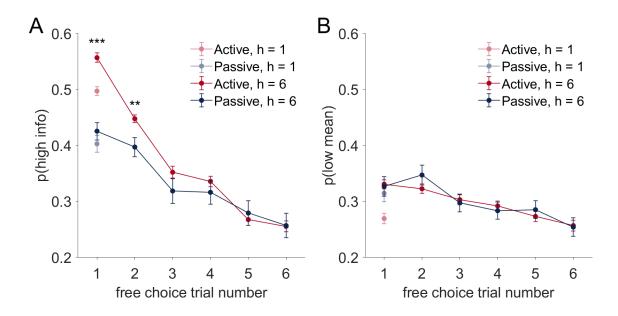
Figure 4: Model-free measures of directed (A) and random (B) exploration over the course of the games in Experiment 1. The difference in directed exploration (A) between active and passive conditions persists into the second trial of the game, where after behavior of the two groups is similar. Conversely for random exploration, the only difference is the increased variability in horizon 1 in the passive condition — behavior in the horizon 6 games is almost identical in the two groups.

Overall performance

Overall performance for the active and passive conditions is shown in Figure 5. This pattern of behavior is very similar to the pattern we had observed in Experiment 1 in the between-subjects design (Figure 2).

For accuracy, there was no difference between the active and passive groups in horizon 1 games (t $(67) = .48$; $p = .636$). In the horizon 6 games, an ANOVA with trials and active-passive condition as within-subject factors showed a main effect of trials ($F(5, 737) = 22.96$; $p < .001$), no main effect for the active-passive condition ($F(1, 737) = .05$; $p = .818$) and an interaction between two factors ($F(5, 737) = 2.74$; $p = .018$). This interaction was caused by increased accuracy in the passive condition on the first free-choice trial ($t (67) = 2.74$; $p = .008$) that was not carried over to the later trials, as shown in the ANOVA on trials 2-6 in which the interaction between trials and active-passive condition is no more significant: $F(4, 603) = .7$; $p = .594$, while there is still a significant main effect for trials: $F(4, 603) = 22.25$; $p < .001$ and no significant main effect for active-passive condition: $F(1, 603) = 1.41$; $p = .24$.

Reaction times were also similar to those observed in Experiment 1. Here, the resemblance to the reaction time plot in the between-subjects design is almost identical: a gradual reduction in reaction time in the active condition versus its abrupt reduction in the passive condition. The ANOVA analysis with trials and active-passive condition as within-subject factors obtained a significant main effect for trials ($F(5, 737) = 103.31$; $p < .001$) but not for the active-passive condition ($F(1, 737) = 2.22$; $p = .14$) and a significant interaction between them ($F(5, 737) = 13.13$; $p < .001$).

In horizon 1 games, the reaction time for the passive condition was significantly slower than the active condition ($t(67) = 2.53$; $p = .01$). In horizon 6 games, for the trial 1, similar to the between-subjects experiment, in within-subject experiment, reaction times for the passive condition was much slower than the active condition ($t(67) = 3.14$; $p = .002$).

As in experiment 1, in both conditions, participants responded faster over time as they get more familiar with each game; but the changes are smoother in active condition as they gradually respond
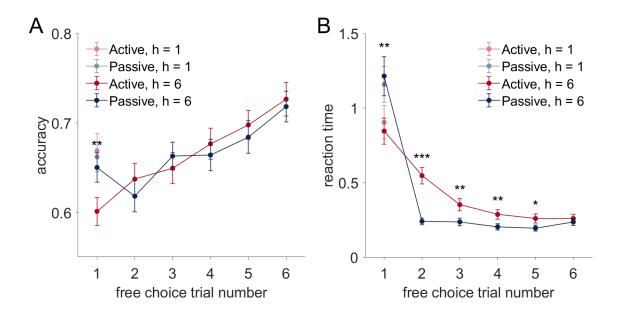
Figure 5: Basic performance in the within-subjects design (Experiment 2). As with the between-subjects design, accuracy (A) increases and reaction time (B) decreases over the course of the horizon 6 games.

faster in trial 2 versus trial 1 and in trial 3 versus trial 2 until they almost respond at a fixed rate in trials 3-6 (similar to the pattern saw in experiment 1). The same similarity goes in passive condition: the participants took about a second to respond in trial 1, but then they responded to the later trials (2-6) much faster and approximately at the same rate.

Directed and random exploration are absent in the passive condition

Like Experiment 1, our main analysis is focused on the first free-choice trial. For both p(high info) and p(low mean), we used a repeated-measures ANOVA with condition (active vs passive), horizon (1 vs 6) and order (active first vs passive first) as three within-subject factors. The results were similar to those we had found in Experiment 1.

For p(high info), the ANOVA showed significant main effects for active vs passive (F(1, 198) = 45.02; p < .001) and horizon (F(1, 198) = 4.05; p = .046) and a significant interaction between horizon and active/passive (F(1, 198) = 3.99; p = .047). We found no significant main (F(1, 198) = 2.36, p = .129) nor interaction effects of order (order*horizon: F(1, 198) = .61; p = .437; order*active-passive: F(1, 198) = .14; p = .704; order*horizon*active-passive: F(1, 198) = .11; p = .740). Post-hoc t-tests showed that the main effect of active-passive condition was due to a reduction in p(high info) in the passive condition in both horizon conditions (for horizon 1: t(67) = 3.55; p < .001; for horizon 6: t(67) = 5.67; p < .001). Similarly, the interaction effect was due to a decrease in the change in p(high info) with horizon (t-test on change in p(high info) between horizon 1 and horizon 6: active condition, t(67) = 2.69, p = .009; passive condition t(67) = .002; p = .99).

For p(low mean), the ANOVA showed significant main effect for horizon factor (F(1, 198) = 4.82; p = .03) but not for the active-passive condition factor (F(1, 198) = 1.52; p = .22). In addition we found a significant interaction between the two factors (F(1, 198) = 7.66; p = .006). As with p(high info) there was no main (F(1, 198) = .35; p = .56) nor interaction effects of order (order*horizon: F(1, 198) = .13; p = .72; order*active-passive: F(1, 198) = .52; p = .47; order*horizon*active-passive: F(1, 198) = .0001; p = .99). Post-hoc t-tests showed that this interaction was due to a non-significant increase in p(low mean)

in horizon 1 (t(67) = 1.12; p = .27) and a decrease in $p(\text{low mean})$ in horizon 6 in the passive condition (t(67) = 2.74; p = .008). In line with these results, the horizon-based change in $p(\text{low mean})$ that was present in the active condition (t(67) = 3.45; p < .001) was absent in the passive condition (t(67) = -.42; p = .68).

Taken together with the findings of Experiment 1, results in Experiment 2 suggest that both directed and random exploration are absent in the passive version of the task ((Figure 6)
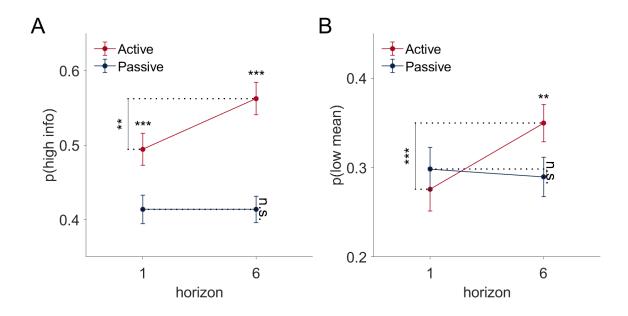

Figure 6: Behavior on the first free-choice trial in the within-subjects design (Experiment 2). As with Experiment 1, both directed (A) and random (B) exploration increase with horizon in the active condition, but not the passive condition.

Correlations between behavior on the active and passive versions of the task

Next we asked whether the different measures of behavior were correlated between passive and active versions of the task. These correlations are shown in Table 1. Each row in this table shows the Pearson correlation coefficient and associated p-value for each of the behavioral measures. As is clear from this table, most measures (accuracy, reaction times, $p(\text{high info})$ in horizon 1 and $p(\text{low mean})$) are highly correlated across active and passive versions of the task. However, the correlations for all variables are stronger in horizon 1 than in horizon 6 games and $p(\text{high info})$ in horizon 6 is not significantly correlated across versions of the task. Most striking, however, is the total lack of correlation between the measures of directed and random exploration, i.e. the change in $p(\text{high info})$ and $p(\text{low mean})$ with horizon. This suggests that directed and random exploration are completely absent in the passive version of the task.

Despite the high correlations of task parameters between two active and passive conditions (except for the $p(\text{high info})$ in h6), the directed and random explorations (which are the difference between $p(\text{high info})$ in two horizons and the difference between $p(\text{low mean})$ in two horizons, respectively) didn't correlate between active and passive conditions at all (low r values of .05 and .03, respectively).

Table 1: Task Parameters Correlations

Task Parameter	r	p
p(high info)_h1	.38	.001
p(high info)_h6	.13	.308
p(low mean)_h1	.66	< .001
p(low mean)_h6	.48	< .001
directed exploration	-.05	.665
random exploration	-.03	.788
accuracy_h1	.68	< .001
accuracy_h6	.38	.002
reaction time_h1	.63	< .001
reaction time_h6	.48	< .001

Behavior on later trials converges between active and passive conditions

Similar to the Experiment 1, in the Experiment 2 we also observe a converging pattern of p(high info) and p(low mean) in the active and passive conditions through the trials in horizon 6 games (Figure 7). As indicated in the Figure, p(high info) and p(low mean) in the active condition are significantly higher than in the passive condition in trial 1. However, for the later trials, the difference in p(high info) and p(low mean) between active and passive conditions are no more significant and they getting close to each other, except for the p(high info) in trial 4.
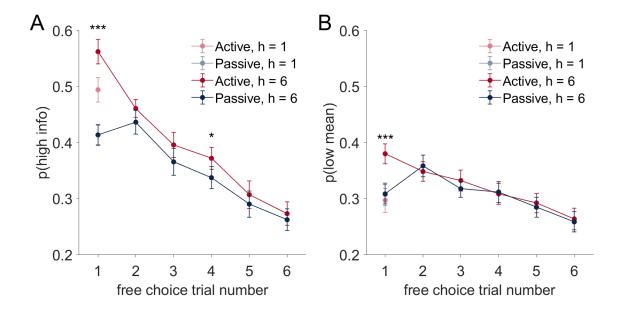

Figure 7: Measures of directed (A) and random (B) exploration over the course of the games in Experiment 2. Unlike the between-subjects design, behavior in the active and passive conditions quickly converges after the first free choice trial for both directed and random exploration.

Discussion

In this paper, we took a closer look at a failed replication from our lab. In particular, we examined our failure to detect evidence for horizon-dependent changes in directed and random exploration in a passive version of the Horizon Task. Because we had previously found this behavioral effect in multiple studies using an active version of the task (Wilson et al., 2014; Zajkowski et al., 2017; Somerville et al., 2017; Warren et al., 2017), we hypothesized that this failure to replicate may have been due to the change between active and passive versions of the task. In both a between-subjects (Experiment 1) and within-subjects (Experiment 2) design we found that participants showed increased ambiguity aversion and reduced horizon-dependent directed and random exploration in the passive version of the task. This clearly demonstrates how behavior in the Horizon Task is critically dependent on the nature of the forced-choice trials, and that our failed replication was caused by the seemingly minor shift from active to passive versions of the task.

With regard to explore-exploit behavior, the obvious question is why do we see such profound differences in behavior between the passive and active versions of the task? Why exactly are people more ambiguity averse in the passive condition? And why do the horizon effects on directed and random exploration disappear?

In some sense, perhaps we should not have been surprised to see some differences between the active and passive versions of the task. In the education literature, it is widely accepted that active learning, whereby the learner controls the input of information, is more effective than passive learning (Kolb, 2012; Kuhn, Black, Keselman, & Kaplan, 2000; Bruner, 1961; Boekaerts, 1997). More recently, in psychology, Gureckis and colleagues have demonstrated improved learning and memory in active vs passive versions of a number of tasks (Gureckis & Markant, 2009; Markant, Ruggeri, Gureckis, & Xu, 2016; Markant & Gureckis, 2014; Markant & Gureckis, 2010), reviewed in (Gureckis & Markant, 2012). Indeed, worse learning in the passive case may explain the reduced performance in the passive condition - at least in the horizon 1 games in Experiment 1 where $p(\text{low mean})$ is increased relative to the active condition (Figures 4B. However, such an effect on learning does not obviously explain the other changes in behavior such as increased ambiguity aversion and reduced horizon effect for both directed and random exploration. For ambiguity aversion, our results are reminiscent of the difference in risk taking behavior that occurs between described and experienced versions of the same gambles (Hertwig, Barron, Weber, & Erev, 2004). In particular, when choosing between gambles that are described (e.g. a 50% chance of $100 vs a 100% chance of $4) participants are risk averse for gains and risk seeking for losses (Tversky & Kahneman, 1973). Conversely, however, when the same gambles are not explicitly described but learned from *experience* (via sampling of hypothetical outcomes) then this pattern reverses and participants are risk seeking for gains and risk averse for losses. Equating the experienced case with the active condition and the described case with the passive condition, could it be that a difference between description (passive) and experience (active) is driving the reduced information seeking in horizon 1 via risk aversion? More work will be needed to test whether this is the case. Of particular interest would be the case of losses, where, if this hypothesis is correct, we would expect to see *increased* information seeking in the passive condition.

For the reduced horizon effect in the passive condition, perhaps the simplest explanation is a difference in attention between active and passive versions of the task. In the active version, information arrives sequentially and predictably, immediately after the participants press the button for the forced trial. Thus, participants are able to allocate attention to each reward in turn and have plenty of time to consider how the horizon condition (which is presented implicitly in the size of the bandits Figure 1C) affects their decision. In contrast, in the passive condition, information from the forced-choice trials arrives all at once after a short time delay from the previous game. In this case, participants must (at least implicitly) decide how to allocate attention between all of this different information, causing them to fail to include the horizon as part of their task set for the decision. Clearly more work manipulating the timing of the passive task (for example using sequentially, but passively presented reward) will be necessary to test this hypothesis.

With regard to the replication crisis, our results represent a single data point in which a failed replication can be 'explained away' by a seemingly minor change in task design. While it is obviously important not to overinterpret a single data point (as anyone who has followed the replication crisis knows only too well!) our finding does at least offer an existence proof whereby a failed replication can be accounted for by a moderator variable like this. More generally, to the best of our knowledge, the prevalence of failed replications that can be explained away like this is unknown. While Cesario (2014) lists some examples where moderator variables could, in theory, hinder replication, we are not aware of a systematic accounting of moderator-explained failed replications. Clearly such a list would help determine whether moderator effects are a major cause of failed replications or simply a theoretical possibility that arises only in rare cases like ours.

Methods

Participants

Experiment 1: Between-subjects experiments

We collected data from a total of 521 participants (ages 18-31, average = 19.33; Females = 278, Males = 243) who took part in various versions of the Horizon Task in our lab between 2015 and 2016. Participants were recruited through the Psychology subject pool at the University of Arizona and received course credit for their participation. The experiments were not explicitly designed to evaluate the effects of active exploration versus passive exploration. However, in some of them (229 participants) the passive version of Horizon Task was used while in others (292 participants), the active version was used. All participants gave informed consent and the study was approved by the Institutional Review Board at the University of Arizona.

Experiment 2: Within-subjects experiment

89 participants were recruited from the Psychology subject pool at the University of Arizona and were compensated by receiving course credits. Of those 89 remaining participants, 68 participants completed both conditions and were included in the final analysis (ages 18-22, average = 18.37; Females = 41, Males = 27). All participants gave informed consent and the study was approved by the Institutional Review Board at the University of Arizona.

Behavioral Task

Horizon Task

In order to study explore-exploit dilemma in the narrow and controlled environment of the lab, we used a simple version of the bandit task. Bandit task refers to a category of tasks that basically involve choosing between different slot machines with different patterns of payout.

The term "Horizon" refers to a key feature in the task which is the number of trials the subject is given in a given situation. In the short horizon, you have only one shot (horizon 1); you choose between a better-known option and a riskier one, then you receive a reward based on your choice, and this is the end of the game (you will proceed to the next game). In the long horizon, you have more opportunities to choose between those options in the same game. For example, in a horizon 6 problem, you have six shots to choose between the left or right bandit. If you choose one of them and it turned out to be very bad, you have more chances to correct yourself and further modify your future decisions.

In each game, subjects see two bandits, either each with five cells (for horizon 1 condition), or each with ten cells (for horizon-6 condition) as shown in Figure 1C. Each pair of cells (one on the left bandit and other on the right bandit) represents a trial. The first four trials of each game have already been played and each contains a number on one of the bandits (the chosen one), representing the obtained

reward, and an "XX" on the other bandit which wasn't selected. The other remaining trials (one for horizon 1 condition and six for horizon 6 condition) is left blank for the subjects to uncover.

By changing the number of unveiled cells in one side (which dictates the other side) in the exemplary trials, we manipulate the amount of information available for each bandit. There are three conditions in this regard:

- 3 reward examples for the left bandit and 1 for the right bandit: [3 1] condition,

- 2 reward examples for each bandit: [2 2] condition, and

- 1 reward example for the left bandit and 3 for the right bandit: [1 3] condition.

By doing this, we can define the uncertainty based on the amount of available information. So, there are two uncertainty conditions:

- [2 2] condition in which there is no difference between two bandits regarding uncertainty, and

- [1 3] condition in which there is more information available on one of the bandits than the other.

The dependent variable in our experiment is the subject's choice on their first free trial (the trial 5 if counting the exemplary ones). Another manipulation in our task is the difference between the mean of rewards between two bandits. After each trial, the reward is sampled from a Gaussian distribution with a fixed standard deviation of 8 points. Each reward is between 1 and 100 points. The generative means of the underlying Gaussians are different for the two bandits and remains stable within a game. In each game, the mean of one bandit is set to either 40 or 60 points and the mean of the other bandit is set in a way that the difference between the two bandits' means is either 4, 8, 12, 20, or 30.

Participants are instructed in the task by the use of a set of illustrated onscreen instructions. It is explicitly conveyed that the means of the two bandits are constant over a game and that the variability of rewards is constant over the whole experiment. Participants are told to try to maximize their earned points.

The two information conditions in the horizon task ([1 3] vs. [2 2] condition) would be used to quantify directed and random exploration. Directed exploration is quantified as the probability of choosing the high information option in the [1 3] condition, p(high info), and random exploration is quantified as the probability of choosing the low mean reward option in the [2-2] condition, p(low mean).

Active vs. Passive Horizon Task

The only difference between Active and Passive versions of the horizon task is in the way the exemplary trials are presented. In the Passive version, the four example trials are shown at the same time. Participants see all the available information for both bandits at once and could make their own decision on the trial 5 (the first free trial) immediately after (Figure 1B)

In the Active version, the example trials are shown one-by-one and right after a motor response from the participant (Figure 1A). The motor response required to proceed the task is simply hitting on the left or right arrow. The first thing participants see is two bandits with blank and clear cells except for one cell; the first cell on one of the bandits would be colored which demands participants to hit the according left or right arrow key. After the appropriate response is given, the example reward points would appear on the selected bandit and an "XX" mark would appear on the according cell in another bandit. This procedure would be repeated for 3 remaining exemplary trials as well. After these four example trials, both fifth cells on two bandits become green-colored which means that the participant is now free to choose between right and left bandits.

In Experiment 1, Active vs Passive was varied between subjects while in Experiment 2, Active vs Passive was varied within subjects with order counterbalanced. Each participant plays 160 games in two parts (80 games for each part). Based on the subject ID, we counterbalance the order of conditions.

References

Boekaerts, M. (1997, 6). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. doi: 10.1016/S0959-4752(96)00015-1

Bruner, J. (1961). *The Act of Discovery* (Vol. 31).

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2018, 9). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. doi: 10.1038/s41562-018-0399-z

Cesario, J. (2014, 1). Priming, Replication, and the Hardest Science. *Perspectives on Psychological Science*, *9*(1), 40–48. doi: 10.1177/1745691613513470

Earp, B. D., & Trafimow, D. (2015, 5). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 621. doi: 10.3389/fpsyg.2015.00621

Ferguson, C. J. (2015, 9). "Everybody knows psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist*, *70*(6), 527–542. doi: /10.1037/a0039405

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016, 3). Comment on "Estimating the reproducibility of psychological science. *Science*, *351*(6277), 1037–1037.

Gureckis, T. M., & Markant, D. B. (2009). Active Learning Strategies in a Spatial Concept Learning Game. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, *31*(31), 3145–3150.

Gureckis, T. M., & Markant, D. B. (2012, 9). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, *7*(5), 464–481. doi: 10.1177/1745691612454304

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004, 8). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. doi: 10.1111/j.0956-7976.2004.00715.x

Kolb, D. A. (2012). *Experiential learning : experience as the source of learning and development*.

Krueger, P. M., Wilson, R. C., & Cohen, J. D. (2017). Strategies for exploration in the domain of losses. *Judgment and Decision Making*, *12*(2), 104–117.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495–523. doi: 10.1207/S1532690XCI1804{_}3

Markant, D. B., & Gureckis, T. M. (2010). Category Learning Through Active Sampling. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, *32*(32), 248–253. doi: 10.1.1.174.7434

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122. doi: 10.1037/a0032108

Markant, D. B., Ruggeri, A., Gureckis, T. M., & Xu, F. (2016). *Enhanced Memory as a Common Effect of Active Learning* (Vol. 10) (No. 3). doi: 10.1111/mbe.12117

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015, 9). Is psychology suffering from a replication crisis?: What does 'failure to replicate' really mean? *American Psychologist*, *70*(6), 487–498. doi: 10.1037/a0039400

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ...

Yarkoni, T. (2015, 6). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. doi: 10.1126/science.aab2374

Pashler, H., Coburn, N., & Harris, C. R. (2012, 8). Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments. *PLoS ONE*, *7*(8), e42510. doi: 10.1371/journal.pone.0042510

Pashler, H., & Harris, C. R. (2012, 11). Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science*, *7*(6), 531–536. doi: 10.1177/1745691612463401

Sayre, F., & Riegelman, A. (2018). The Reproducibility Crisis and Academic Libraries. *College & Research Libraries*, *79*(1), 2–9. doi: 10.5860/crl.79.1.2

Simons, D. J. (2014, 1). The Value of Direct Replication. *Perspectives on Psychological Science*, *9*(1), 76–80. doi: 10.1177/1745691613514755

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017, 2). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of experimental psychology. General*, *146*(2), 155–164. doi: 10.1037/xge0000250

Stroebe, W., & Strack, F. (2014, 1). The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science*, *9*(1), 59–71. doi: 10.1177/1745691613514450

Tversky, A., & Kahneman, D. (1973, 9). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. doi: 10.1016/0010-0285(73)90033-9

Warren, C. M., Wilson, R. C., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., Cohen, J. D., & Nieuwenhuis, S. (2017, 4). The effect of atomoxetine on random and directed exploration in humans. *PLOS ONE*, *12*(4), e0176034. doi: 10.1371/journal.pone.0176034

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014, 12). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of experimental psychology. General*, *143*(6), 2074–81. doi: 10.1037/a0038199

xkcd: Negative Results. (n.d.). Retrieved from `https://xkcd.com/2020/`

Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *eLife*, *6*. doi: 10.7554/eLife.27430