Possibility Generation Study

Dataset:

Thus far, three separate studies compose the data we are utilizing. First, there is the original possibility generation study and its replicated version. The former study lacked measures about possibility goodness ratings, possibility generation length (in time), openness, and PANAS measures, whereas these were present in the replicated study. In addition, a third, slightly different possibility generation study is present within the data. In this study, participants are similarly shown several vignettes, but asked to generate eight possibilities per vignette instead of six.

Study	**Timed Generations?**	**Generation Ratings**	**Reflective Measures**
Study A	No	No	BDI, BAI
Study A’	Yes	Yes	BDI, BAI, PANAS, Openness
Study B	Yes	Yes	None

Statistically Significant Results:

Test	**Replicated**	**Further Studies Needed**
Average sentiment per possibility number	**Yes**	
Sentiment gradient vs. beck score	No	
Sentiment gradient vs. openness score	-	Yes
Global average sentiment vs. total time	No	
Semantic evaluation difference (SED) vs. beck	-	Yes
SED vs. PANAS	-	Yes
SED vs. openness	-	Yes
Reflection gradient vs. total time	-	Yes
Reflection gradient vs. sentiment gradient	-	Yes
Reflection gradient vs. time gradient	-	Yes
Successive semantic distance vs. time	**Yes**	
Semantic exploration vs. global average sentiment	**Yes**	
Semantic exploration vs. total time	**Yes**	
Generation time vs. semantic distance from the center of mass (SDCOM)	**Yes**	
Total time vs. possibility number	**Yes**	
Semantic space similarity vs. possibility number	**Yes**	
Trajectories Clustering using KMeans and HDBSCAN	?	
Semantic distance vs. beck	No	
Total semantic exploration vs. possibility number	**Yes**	

Replicated Results:

Test
Average sentiment per possibility number
Successive semantic distance vs. time
Semantic exploration vs. global average sentiment
Semantic exploration vs. total time
Generation time vs. semantic distance from the center of mass (SDCOM)
Trajectories Clustering using KMeans and HDBSCAN
Total semantic exploration vs. possibility number
Total time vs. possibility number
Semantic space similarity vs. possibility number

Average sentiment per possibility number:

Summary:

Suppose every generation takes the form $(text,\ order)$, where text is the content of the generation, and order is the relative order of that generation within the vignette trial (e.g., 1st, 2nd, 3rd, etc.…) This test sums all generations for a particular possibility number (i.e., 3rd) and discovers the average sentiment value.

Initial: (Study A)

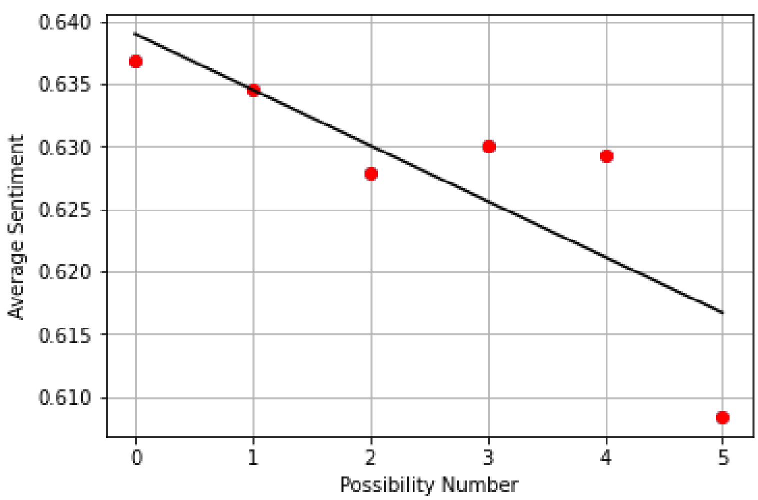


Figure 1

```
r-value:-0.824304
p-value:0.043592
```

Replicated: (Study A')

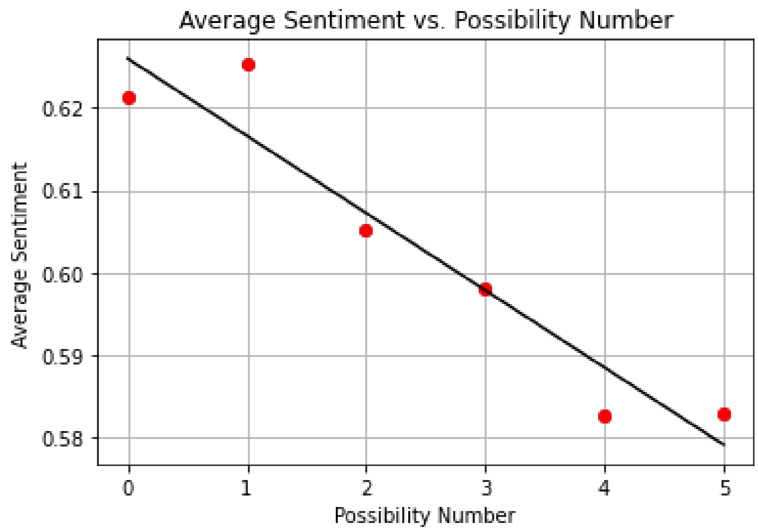


Figure 2

```
r-value:-0.954077
p-value:0.003115
```

Successive semantic distance vs. time

Summary:

Given that generations have a particular order, we wondered about the relationship between successive generations across semantic and temporal space. So, this test finds the semantic distance between all ith and $i + 1$th generations and plots this against the time that elapsed between these generations (or the amount of time it took to generate the i+1th possibility).

Initial: (Study B)

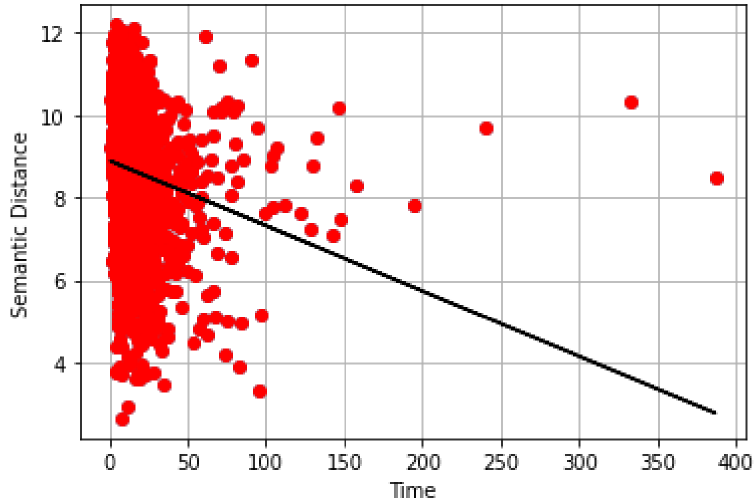


Figure 3

```
r-value:-0.180382
p-value:0.000000
```

Replicated: (Study A')

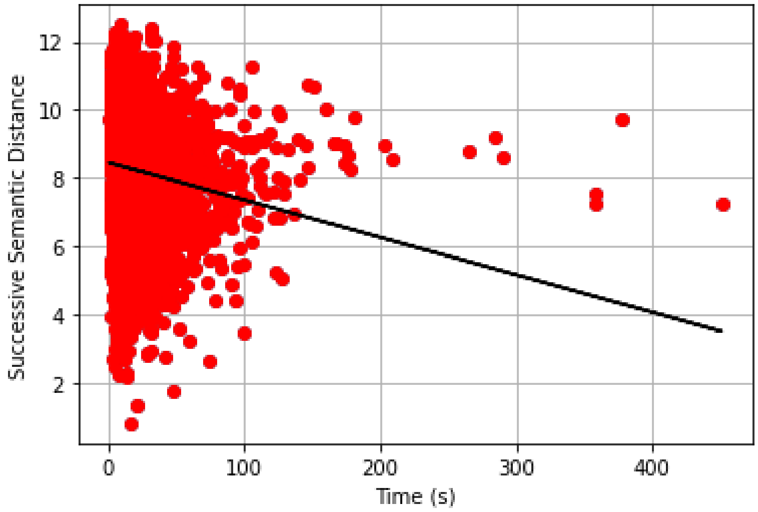


Figure 4

```
r-value:-0.135768
p-value:0.000000
```

Semantic exploration vs. global average sentiment

Summary:

We define semantic exploration to mean something like: $\sum_{v \in vignettes} \sum_{i=0}^{n} \sum_{j=i}^{n} dist(v[i], j[i])$, where v is a particular vignette, n is the number of generations per each vignette (i.e., 6), $v[x]$ is the xth generation of the vth vignette, and $dist(x, y)$ is the Euclidian distance between these two generations in semantic space.

Initial: (Study A)

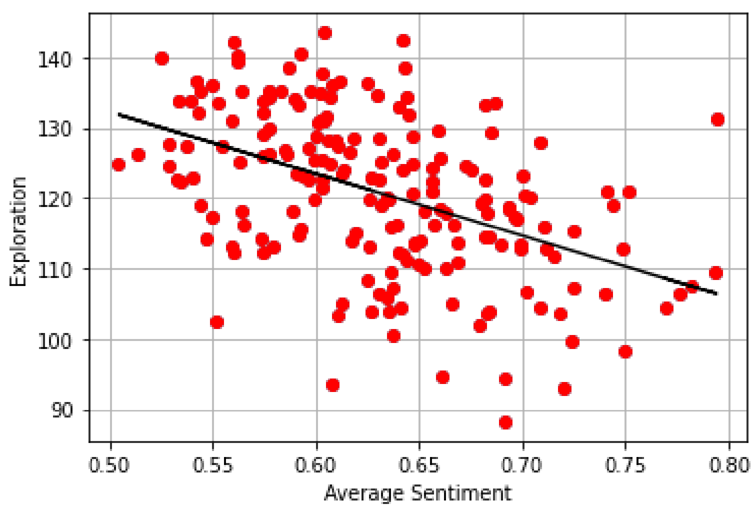


Figure 5

```
r-value:-0.468150
p-value:0.000000
```

Replicated: (Study A')

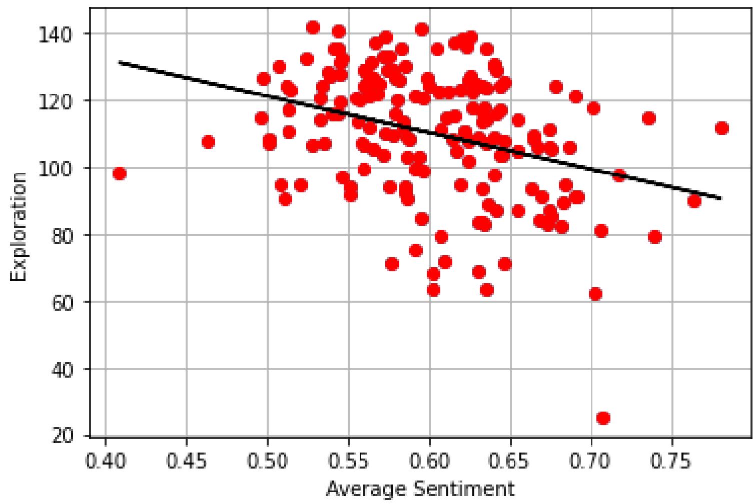


Figure 6

```
r-value:-0.332320
p-value:0.000006
```

Semantic exploration vs. total time

Summary:

Here, we conceptualize semantic exploration slightly differently. We define semantic exploration to mean: $\sum_{v \in vignettes} max_{i \in n, j \neq i \in n} dist(v[i], j[i])$. That is, the sum of the max distances between intravignette generations across all vignettes.

Initial: (Study B)

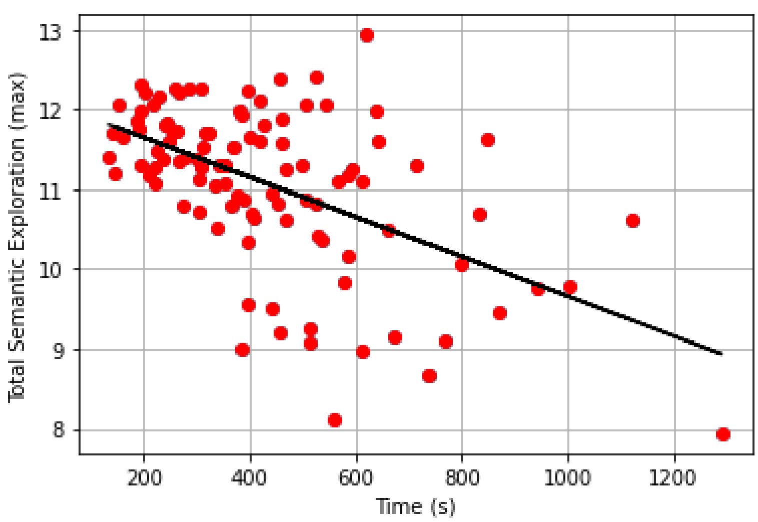


Figure 7

```
r-value:-0.535968
p-value:0.000000
```

Replicated: (Study A')

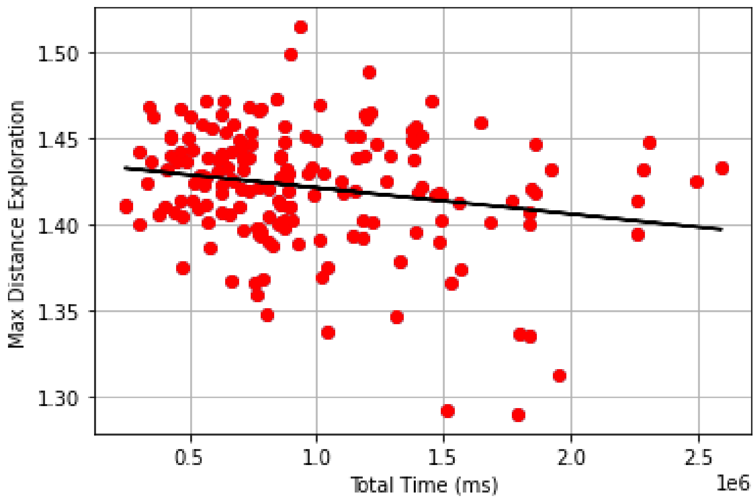


Figure 8

```
r-value:-0.211152
p-value:0.004666
```

Generation time vs. semantic distance from the center of mass (SDCOM)

Summary:

As participants generate possibilities in each vignette, we can conceive of them building something like a semantic center of mass. Each additional possibility they generate will shift this footprint – the magnitude of this shift is contingent on the distance of the novel possibility from the existing semantic center of mass. More formally, if *j* possibilities have been generated so far, we can conceive of the center of mass as the index-wise average across all the corresponding embedding vectors. The distance between the novel vector, and this existing vector, is what is being plotted here against time.

Initial: (Study B)

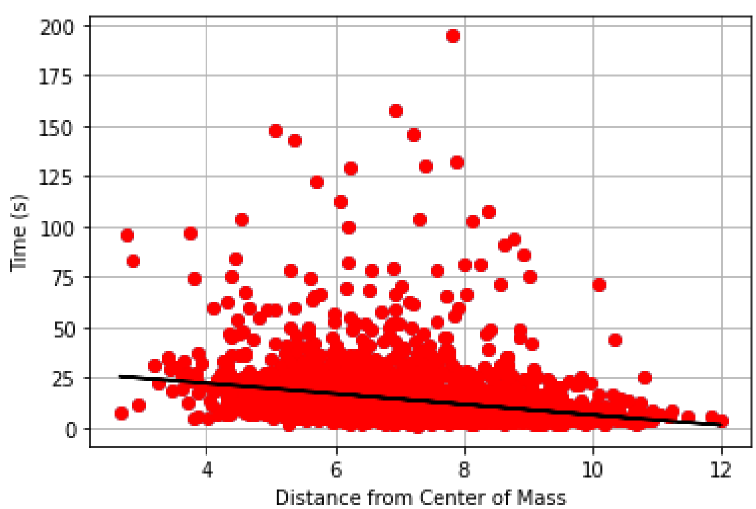


Figure 9

```
r-value:-0.251934
p-value:0.000000
```

Replicated: (Study A')

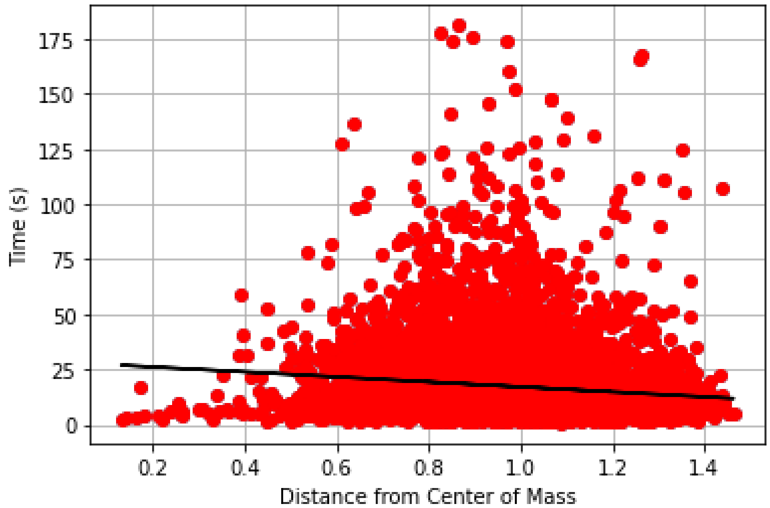


Figure 10

```
r-value:-0.122051
p-value:0.000000
```

Total semantic exploration vs. possibility number

Summary:

The semantic distance defined in this test is identical to the metric described above. For each possibility number, there will be some *n* generations. Hence, we find all possible $\binom{n}{2}$ pairs of embeddings, summing the distance between these embeddings for each possibility number.

$$\sum_{p \in possibility\ number} \sum_{i=0}^{n} \sum_{j=i}^{n} dist(p[i], p[j]).$$

Initial: (Study A)

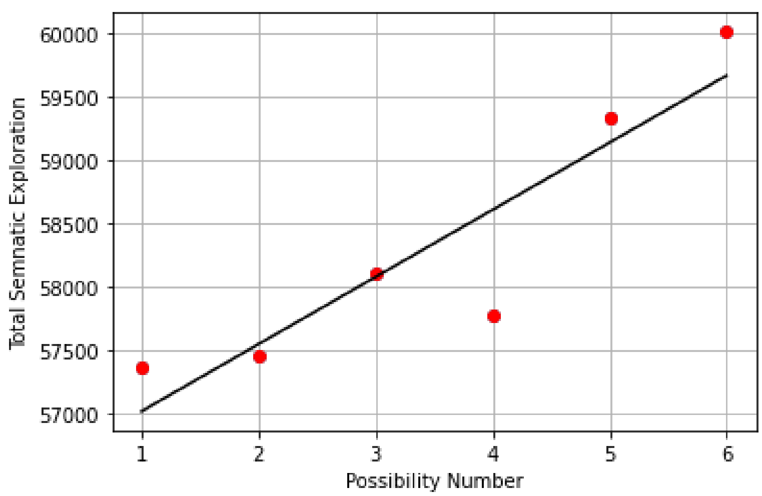


Figure 11

```
r-value:0.911951
p-value:0.011288
```

Replicated: (Study A')

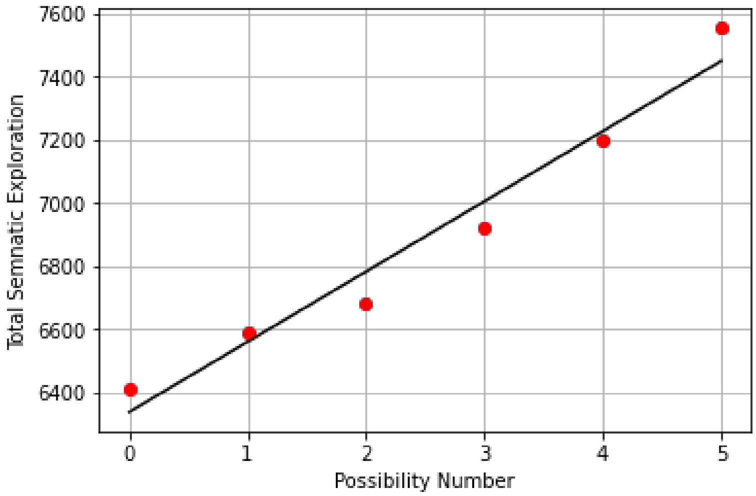
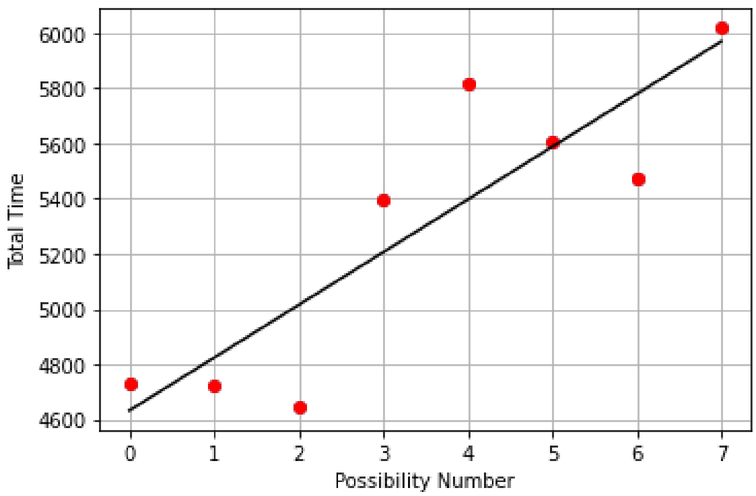


Figure 12

```
r-value:0.980371
p-value:0.000574
```

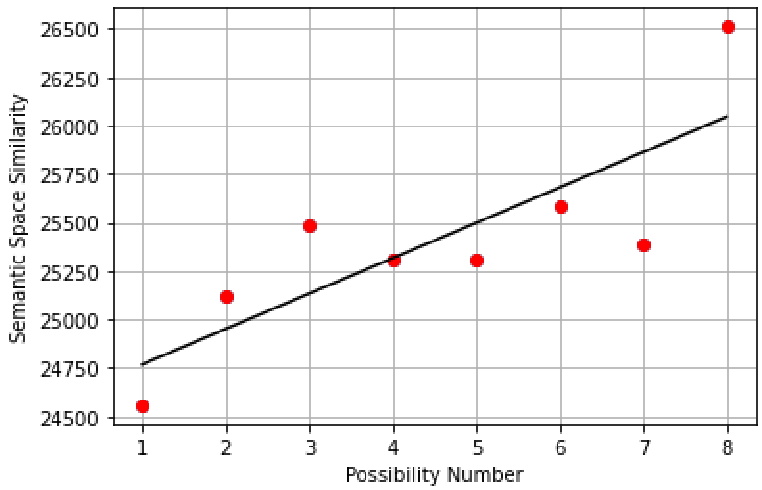
Total time vs. possibility number

Initial: (Study B)




```
r-value:0.875846
p-value:0.004350
```

Replicated: (Study A')




```
r-value:0.818679
p-value:0.012950
```

Semantic space similarity vs. possibility number

Summary:

Whereas semantic exploration measures the extent to which participants explore semantic space through their generations, semantic similarity measures the extent to which participants explore similar regions of a shared semantic space. Formally, this likeness is measured by first creating a localizing vector. A localizing vector is defined as follows:

$$localizing\ vector\ =\ average(for_{p \in P} embedding(p)).$$

A localizing vector for the 0^{th} generation is the average vector derived from all 0^{th} generations the participant created throughout the study. From there, semantic space similarity measures the distance between every participant's localizing vector. Specifically,

$$Semantic\ space\ similarity_i = \sum_{p\ \in P_{localizing,i}} \sum_{j \neq p\ \in\ localizing_i} dist\ (embedding(p), embedding(j)).$$

Initial: (Study B)

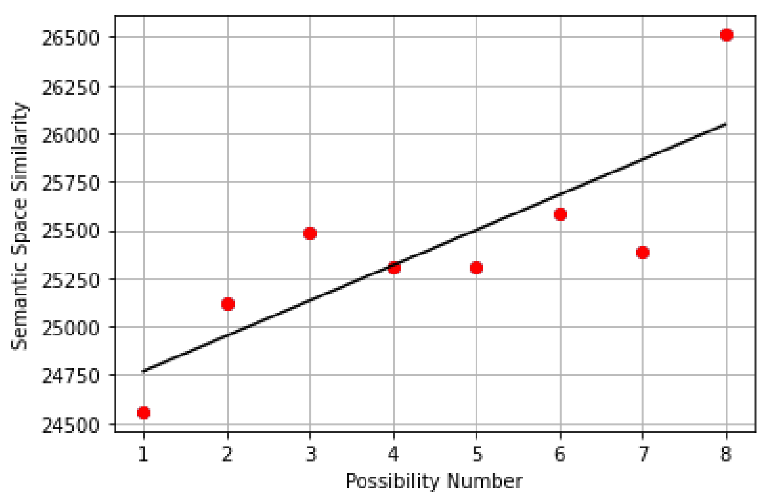


Figure 13

```
r-value:0.818679
```

p-value:0.012950

Replicated: (Study A')

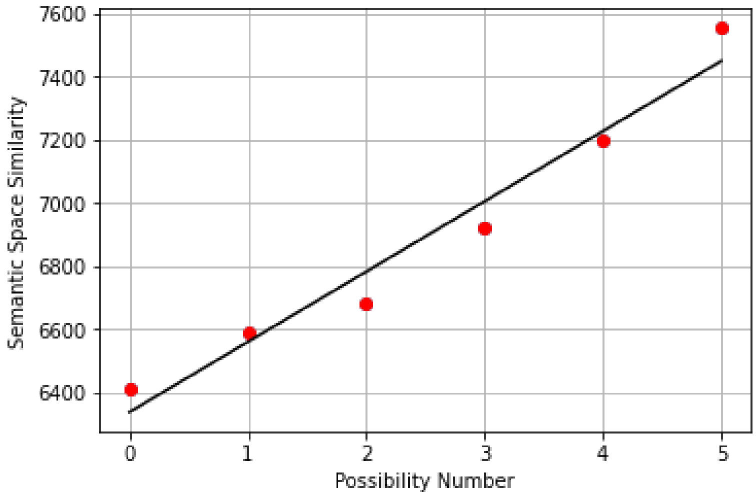


Figure 14

r-value:0.980371
p-value:0.000574

Discussion

Two main themes seem especially prominent in the data.

1. *There is an inverse relationship between semantic exploration and time*

Be it successive semantic distance, total semantic exploration, or semantic distance from the center of mass, all of these metrics, when plotted against time, paint a strikingly similar picture: the longer participants take to form generations, the smaller their domain of semantic exploration. Two prominent forms of explanation come to mind in light of this trend. We might think that the longer time intervals in specific tasks reflect the involvement of the more computationally intensive, model-based processing system. This system tends to explore a far narrower range of possibilities than its model-free, computationally cheap counterpart. Alternately, we may explain this result as a consequence of generation fatigue. That is, it is *because* participants cannot come up with appropriate generations that they tend to take longer, and these generations are less original.

2. *There is the fingerprint of a dual systems theory of modal cognition*

First, observe that as average sentiment declines through the generation task (fig. 1&2.) Second, semantic exploration is inversely related to global average sentiment (fig. 5&6.) Third, semantic exploration increases through the task (fig., 11&12.) Fourth, semantic space similarity decreases through the task (fig. 13&14.) Together, these paint a compelling picture in favor of a dual system theory of modal cognition. Namely, participants tend to begin the task relying on a computationally cheap, cached value type of cognition to generate possibilities before transitioning to more dynamic processing. Participants tend to explore less early on - relying on consideration sets that tend to be shared across the participant space and have historically high expected value (further evinced by their higher sentiment scores.) As participants deviate from this cognition later in the task and begin to rely more on dynamic processing, we see semantic exploration increase commensurately. Predictably, this online decision-making, while task-specific, adds greater variability to the mix – perhaps explaining why average sentiment scores decline throughout the task.