Introduction

In the following document, I propose changes to the design of our experiment to detect inter-task effects.

Design Considerations

- **permutation of test images** In our previous experiment, we attempt to assert that priming effects "wash-out" at a partucular rate. Unforturately, since different images have different content, one cannot expect them to have the same *susceptibility* to priming.

  To solve that, we should permute the images in the test set. But how?

  We discussed two ways to do this. My personal feeling is still that a simple rotation (the "5x5 set-up") would be best:

  ```
  1 2 3 4 5
  2 3 4 5 1
  3 4 5 1 2
  4 5 1 2 3
  5 1 2 3 4
  ```

  But an alternative, as we discussed is an exhaustive permutation, using a smaller test set (the "3x6 set-up"):

  ```
  1 2 3
  1 3 2
  2 1 3
  2 3 1
  3 1 2
  3 2 1
  ```

  I prefer the 5x5 set-up for two reasons:

  1. We can watch how priming falls off for a given image A when A occupies 5 positions. This is long enough to give a good picture of the "washout curve".
  2. We can present 5 independant "washout curves", (rather than 3). By following each of the 5 images through each of the 5 positions.

  With the 3x6 set-up, we only get three washout curves, each of length 3.

  What follows is one possible critique that we may meet if we take the 5x5 set-up. I believe that this critique, although reasonable *prima facie*, is logically flawed:

  Possible critique: "In your experiment, image 4 always follows image 3—doesn't that potentially influence turker's responses to image 4?"

  Rebuttal: "Of course, image 3 certainly influences turkers responses to image 4. But that's true both when the turker was primed with cultural images as well as when primed with food images. Any differences that we observe for the same ordering of test images but for different primes are certainly caused by the difference of the primes. Since we showed similar priming washout dynamics for 5 different images, we are very confident that they reflect general inter-task dynamics. Repeating the experiment for all permutations would add 23 replicates to our 5, and at considerable cost. The results are clear already with 5 replicates of washout profiles."

- **Priming target**. Untill now I've still been referring to turkers primed by **food** images and **culture** images. I think that we should change our priming targets to **food** and ***artifacts***. Artifacts are more

"imageable", and so I think easier to prime. Also, I think its easier to defend our classification of labels as being 'food' vs 'artifact' when building the ontology, than is the case for 'food' vs 'culture'.

- **Don't use "cross-treatments".** We have treatments based on *framing*, in order to compare the effect-size of inter-task priming to something that's been studied already. Let's keep the framing treatments "pure" in the following sense: don't provide *any* initial images for the framing treaments. I.e. turkers that do the framing treatments go straight to labelling the test set. This is related to the next bullet.

- **Don't include a "neutral" initial image set.** Before we had the initial image set that we called *ambiguous*. I think we should only have *food* and *artifacts* initial image sets. They can be compared together for a direct read on inter-task effects. In my opinion, there really is no such thing as a "neutral" set of pre-tasks, and including one just makes analysis more complicated and less direct. Meanwhile, since there is no "neutral" initial tasks, our framing treatments don't have *any* initial tasks, which means that all effects are the result of the framing, and that framing occurs *immediately* before labeling the test images, which is better design in my opinion.

- **What kind of "framing" should we do?** The critque that we got for the framing from the last experiment was that it is too weak. People hypothesized that most turkers didn't even read the funder name. One way to amp it up, is to make turkers chose the funder name from a drop-down list. I think that makes the most sense.

  Recently I was considering the possibility of also changing the language, to something like "The purpose of this study is to understand people's perceptions of tools and objects." vs "... people's perceptions of food and ingredients." That would very likely be *much* stronger. This might mean we wouldn't "beat" it with inter-task effects.

  Theoretically, there is not a huge difference, but practically, I think it's more useful in the context of a publication, to say "X is stronger than weak Y" than "X is almost as strong as strong Y".

  Consider, however, that the framing treatments are much *cheaper* than the inter-task treatments. When we do the inter-task treatments, each priming treatment (food, artifacts) has to be done 5 or 6 different ways (sub-treatments) because of the desire to achieve independance of ordering. But for the framing treatments, I see no reason to do that. This means that for the cost of about 300 more HITs, we could actually do both the strong and weak framing. If we beat strong framing, we could report it. If we only beat weak framing, then report *that*. Is that ethical? My gut tells me that is totally fine. People generally don't report unremarkable aspects of their results. True this makes blind spots in the community, but that needs to be fixed at a level at which we are not operating anyway.

- **Use stock photos.** I would be much more comfortable using stock photos. For just $1 we can purchase the rights to display them on a public server. Otherwise, I'm worried, not that the copyright owners will come after us (doubtful), but that the journal will be uneasy about being associated to work that technically breached copyright (possible). I selected a bunch of images based on the priming targets *food* and *artifacts*. They're on the last page of this doc.

- **Money.** Based on costs last time, if we use the "5x5 set-up" the total cost is $C = 1950 * 0.5 * (1.1) =$ $1072.

  Using the 3x6 set-up, means we have an extra permutation for both the food and artifacts inter-task treatments, so 2 extra treatments, and the cost is slightly more: $12650. This is the same cost as if we used the 5x5 set-up, and include both a strong and weak version of the framing.

  Finally, If we do the 3x6 set-up with both strong and weak framing we have 16 treatments and the cost is $1430.

  The original experiment was much cheaper because we didn't have multiple permutations: $495.

Images Sets

Artifacts priming set

- www.dollarphotoclub.com/68948332
- www.dollarphotoclub.com/67422414
- www.dollarphotoclub.com/68305501
- www.dollarphotoclub.com/65973300
- www.dollarphotoclub.com/61117158

Food priming set

- http://www.dollarphotoclub.com/65776429
- http://www.dollarphotoclub.com/68630458
- http://www.dollarphotoclub.com/60339399
- http://www.dollarphotoclub.com/50058557
- http://www.dollarphotoclub.com/49690807

Test set

- http://www.dollarphotoclub.com/58430347
- http://www.dollarphotoclub.com/55333951
- http://www.dollarphotoclub.com/67985597
- http://www.dollarphotoclub.com/47899888
- http://www.dollarphotoclub.com/59285361