Goals

This document presents my revised plan for the priming experiment. First, let me present the goals of the experiment, in the form of questions that I intend to answer:

1. How can we characterize and quantify HPU hysteresis? (this actually theoretical rather than experimental)

2. Inter-task effects are a fundamental and difficult-to-avoid kind of hysteresis. How strong are inter-task effects for the canonical HPU task of image-labeling?

3. How do inter-task effects compare to other kinds of priming, like framing?

4. How long do inter-task effects last?

5. Can we use knowledge about inter-task effects to build better HPU algorithms?

Design

The following design should allow us to directly answer, or at least speak to, the above questions. Some auxiliary questions, though, will inevitably get left unanswered.

The basic idea, like before, is to have some treatments that differ in terms of *framing*, and others which differ in terms of the *initial tasks*.

But one change I'd like to propose, is that we take our priming targets to be food and *artifacts*, instead of food and *culture*. So, in one priming treatment, we try to get people to talk about the objects they see, like the statue, the candle, the bottle, the chair, etc, whereas in the other, of course, we try to get people to talk about the food. I have a couple reasons for wanting to switch from culture to artifacts. First of all, I think that its easier (and less contentious) to distinguish food and artifact words while building the ontology. Second, artifacts are more concrete, and I think easier to name, which makes me think that it is an easier priming target.

The treatments plan is in the table below:

treatment code	framing	initial tasks	notes
food-tasks	$\emptyset$	food	use 5 permutations of test-images
artifact-tasks	$\emptyset$	artifacts	use 5 permutations of test-images
food-frame	food	$\emptyset$	use just 1 permutation of test-images
artifact-frame	artifacts	$\emptyset$	use just 1 permutation of test-images

Another difference I'm proposing is that we have no "cross-treatments". What I mean is that the inter-task treatments have no framing statement, and the framing treatments have no initial tasks at all (they go straight to labelling the 5 test images). This is both to save on the total number of unique workers that have to participate, and because this provides a more direct test of framing effects.

I'd like to propose that "framing" be done by including a "purpose statement" So, our framing statements could be: "The purpose of this study is to understand people's perceptions of tools and objects." vs "...people's perceptions of food and ingredients.". This will, almost surely, make framing stronger. But I think that it is a more honest comparison. And it is directly relatable to another experiment in the literature, which investigated framing of that kind.

Like you suggested, people would be asked to choose the "purpose" of the study from a drop down list, so that we know that they read it and so we don't have to deal with spelling mistakes. This can also serve as our "filter" for turkers who are just choosing anything.

I'd also like to propose that we drop the idea of a "neutral" treatment. In the last setup, it might be claimed that AMBG was a neutral treatment, but that claim isn't consistent with how we talk about priming.

More importantly, we don't need AMBG to make our point, and actually I think it just complicates the discussion. I think we should focus on comparing treatments that differ along either a framing or inter-task dimension, and not bother with a "neutral treatment".

Probably the biggest change is to introduce permutations of the test-images. We already talked about doing this—the idea is that in order to address Q 4, we need to test different orderings of the images. I've thought a good deal about exactly how to do the orderings, which is what I want to address next.

Since we have 5 test images, there are 120 permutations. Obviously we won't collect data for all permutations! I think that we need each image to appear in each position of the test-set. That is the only way, it seems to me, that we can test the effect of position without confounding the content of images. Also, it would probably be best to have the first image in the test set be occupied by each of the 5 test images in turn, because it is where we expect the strongest effects.

I think that the best approach is to first generate a "starting permutation" randomly, and then and use the four other orderings that are achieved simply by "rotating" the starting permutation. Like this:

```
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
```

One drawback though, is that the relative ordering is always the same—e.g. image 4 always follows image 3. This means we can't answer an auxiliary question "How do inter-task effects modulate inter-task effects". But it seems to me that trying to answer that is out of scope for this experiment.

This means we would have 12 treatments $(2 * 5 + 2 = 12)$. We have we have five orderings for the first two treatments, which really means that we have 10 treatments dedicated to testing inter-task effects. I don't think we need to do extra permutations for the framing treatments, because that, IMO, is not what we are trying to investigate. It seems to me that this is already a lot of treatments, especially if we want to get over 125 unique workers for each (which I think is a good minimum target—more would be better actually).

If we don't permute the test-images in the framing-based treatments, it does mean that we are unable to answer an auxiliary question: "do different types of priming have different hysteretic dynamics?". I think it's best to leave that aside, otherwise we just have too many treatments.

Images

So, as I mentioned, I was thinking about going in the direction of having food and artifacts be the priming targets.

One thing that I'm concerned about is copyright. It's not that I'm worried that some copyright owner would give us a problem, it's that the journal wouldn't want to print material that is copyrighted. One solution is to use stock photos, which give us the right to host the images on our server for $1 each. Then, in materials and methods, we link to our page that exibits the photos.

Anyway, I think that using stock photos that we actually have the rights to use seems safer. I've looked through the catalogue at http://www.dollarphotoclub.com/, and I found these images:

Artifacts priming set

- http://www.dollarphotoclub.com/55650468
- http://www.dollarphotoclub.com/68491198
- http://www.dollarphotoclub.com/67139257
- http://www.dollarphotoclub.com/65506599
- http://www.dollarphotoclub.com/36887338

Food priming set

- http://www.dollarphotoclub.com/68630458
- http://www.dollarphotoclub.com/65706519
- http://www.dollarphotoclub.com/48722518
- http://www.dollarphotoclub.com/65974609
- http://www.dollarphotoclub.com/44366011

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

Money

This means we'll need about 1950 unique worker-HITs. With 12 treatments, if we want to get 125 unique workers for each, then, we should get about 1950 unique worker-HITs completed. The reason this is more than $125 * 12 = 1500$, is because random assignment of workers to treatments doesn't lead to perfectly balanced treatments. It's *pretty tough* to get balanced treatments because it means having the HITs share state about how many different treatment-types have been completed, and that's not easy in AMT.

This would be a bit more expensive than last time (which cost $495). Assuming we use the same reward as last time, we get a cost of $C = 1950 * 0.5 * (1.1) = \$1072$.

Is that too much? We could drop one of the test-images maybe, of course that means a shorter window to observe priming effects.

Let me know what you think about the whole design. I'm ready to kick off the experiment any time.