**Study Design:**

**Hypotheses:**

1)Given sufficient variation in knowledge, individuals will choose to copy the most copied individual as a proxy of their success compared to proxies unrelated to success

2) People will only choose to copy the most copied when direct access to success information is unavailable.

**Methods**:

*Participants:*

Ten teams of 10 participants for each condition (300 participants altogether) will be recruited via Mechanical Turk. All participants must be above the age of 18. All participants will be given a monetary reward for their time of $10, and will have the opportunity of winning a bonus payment of $20 by achieving over a certain number of points during the task.

*Materials:*

The experimental programme Dallinger is used to generate a game in which teams of players can play and interact at once. 100 questions with two alternative answers each; one correct and one incorrect. The 100 questions are split into four categories of 25 questions each. The categories are named “Geography quiz”, “Weight estimation”, “Language quiz”, and “Art quiz” to participants. An example question for each category is given below:

Geography quiz:

Which City is nearest to Leipzig?

a) Berlin b) Prague

Weight estimations:

How much does an average camel weigh?

a) 480 kg  b) 180 kg

Language quiz:

The word “pisică” means ‘cat’ in which language?

a) Romanian b) Hungarian

Art Quiz:

Who died in 1519?

a) Pablo Picasso b) Leonardo da Vinci

*Procedure:*

100 Participants are given 100 binary choice questions based on four different general-knowledge or ‘trivia’ style categories, 25 in each category. Participants have fifteen seconds to answer each question. If participants don’t know the answer to a question, they are given the opportunity to “Ask Someone Else.” This button allows them to see information about other participants who have answered that question. The information they see depends on the condition, detailed below. They can then choose a participant whose answer they will use for that question. If that participant answered the question correctly, they will also score a point for that question. If they did not, they will not. No one will receive feedback on the right or wrong answers at any point.

Our study comprises of three conditions (A, B and C):

**Condition A: (Irrelevant demonstrator information, copying cues only)**

**Round 1 (first 40 Qs):** Participants will be able to see the Player ID of each participant who answered that question.

**Round 2 (subsequent 60 Qs):**

When choosing to “Ask Someone Else”, participants are asked what type of information about the demonstrators they would rather see, their player ID or the number of times they were copied in Round 1.

**Condition B: (relevant demonstrator information, copying cues only)**

**Round 1:**

Participants will be able to see the individual score of each participant who answered that question.

**Round 2:**

When choosing to “Ask Someone Else”, participants are asked what type of information about the demonstrators they would rather see, their player ID or the number of times they were copied in Round 1.

**Condition C: (relevant demonstrator information, copying and success cues available)**

**Round 1:** Participants will be able to see the individual score of each participant who answered that question.

**Round 2:**

When choosing to “Ask Someone Else”, participants are asked what type of information about the demonstrators they would rather see, their score in Round 1, or the number of times they were copied in Round 1.

**Predictions:**

1. Participants copy the highest scoring participant out of those available in Conditions B & C.
2. Participants copy the most-copied participant out of those available. \*\*Do we use all instances of prestige-based copying here, including any in Round 2 of Condition A (which would be a silly thing to do?) perhaps we keep any condition A prestige copying separate and analyse that separately as a strange form of maladaptive prestige-biased copying?\*\*
3. Participants choose to view prestige info more in Condition B than the other two conditions, because (i) in Condition B copiers can access success info, unlike Condition A where copiers only have access to irrelevant info, and (ii) in Condition B copying info is the only relevant cue available, unlike Condition C where direct success info is available and just as easily accessible.
4. Copying rate is higher in Conditions B & C compared to Condition A because copying is only based on success in Conditions B & C.
5. Participants perform best on the quiz in Condition B & C compared to Condition A because copying is only based on success in Conditions B & C.

***Analyses:***

We will run and compare a series of Bayesian multi-level mixed models using the *Rethinking* package in R and Stan (McElreath 2016).

**Analysis 1:**

To test Predictions 1 and 2, data from all instances of copying when score/prestige information was available/chosen will be used for a multilevel binomial model. The demonstrators will be coded as “highest scoring/copied” or not, to check if participants were consistently choosing the top scorer/most copied out of those that were available for that copying instance. “Highest scorer” or “Most copied” will be the outcome variable. We will use a multilevel model with varying intercepts for group, participant, (and question?)

TopCopy ~ intercept + 1|Participant + 1|group (+1|question?)

presCopy ~ intercept + 1|Participant (+1|question?)

OR:

Chosen.demonstrator ~ interecept + top.scorer + 1|participant + 1|question ?

Chosen.demonstrator ~ interecept + most.copied + 1|participant + 1|question ?

Or, multinomial…??

**Analysis 2:**

To test Prediction 3, a generalised linear mixed model of all data from Round 2 will be used with “info choice” (i.e. prestige or not) as the outcome variable (binomial), with condition as the predictor variables, and a random effect for participant. Condition B will be coded as the baseline, to compare both Condition A and Condition C to Condition B, as we expect more copying to occur in Condition B than either of the other conditions. We will compare this model to a null model.

Wondering whether to have Condition A as the baseline as it is our most “control”- like condition where we wouldn’t really expect them to copy or do well because all the information is useless. I said B initially as that is our strongest prediction, they really should be doing the most prestige copying in this condition and the least in the other two, whereas it’s possible that prestige-bias is so strong that they even use it in Condition A too…

Info.Choice ~ intercept + Condition A + Condition C + 1|Participant + 1|group 1|question?

**Analysis 3:**

To test predictions 4 and 5, a general linear mixed model of all data will be used with overall score/copying rate as the outcome variable and Condition B and C as the predictor variables, with a random effect for participant and group. Condition A will be used as the baseline to compare scores with Condition B and C, as we expect Condition B and C to both have higher scores than Condition A. I have done this as likelihood to choose to copy on any given question by any given person (per trial). I have then used overall score of each person as this accounts for score attained through copying too.

Copying Instance ~ Condition B + Condition C + 1|Participant + 1|group

Overall Score ~ Condition B + Condition C + 1|Participant + 1|group