

THE UNDERESTIMATION OF COMPLIANCE TO DIRECT HELP REQUESTS EXPLAINED BY ADAPTIVE SAMPLING

Danae Arroyos Calvera
Director: Gaël LeMens

Code: W20
Barcelona, June 2013
International Business Economics

TABLE OF CONTENTS

Introduction & Theoretical Background	2
The Model.....	3
Simulation 1. Realistic setup	7
Simulation 2. Neutral setup	7
Simulations 3 and 4. Varying parameter b.....	8
Simulations 5 and 6. Varying sensitivity (parameter s)	9
Simulations 7 and 8. Varying target (“stop”)	9
Simulations 9 and 10. Varying relative attractiveness.....	10
Simulations 11 to 15. Varying tasks difficulty (Prob. of success $rP(i=y)$)	11
Results	13
The Experiment	14
Pretest 1	15
Pretest 2	15
Pilot Experiment 1.....	15
Pilot Experiment 2.....	16
Pre-experiment Survey	17
Experiment 1	17
(General) Discussion and Data Analysis	18
Reference list.....	21
Annex.....	22
Instructions Condition 1	22
Instructions Condition 2.....	23
Post-experiment Survey.....	25

INTRODUCTION & THEORETICAL BACKGROUND

Helping behavior is any intentional behavior that benefits another living being or group (Hogg & Vaughan, 2010). People tend to underestimate the probability that others will comply with their direct requests for help (Flynn & Lake, 2008). This implies that when they need help, they will assess the probability of getting it (De Paulo, 1982, cited in Flynn & Lake, 2008) and then they will tend to estimate one that is actually lower than the real chance, so they may not even consider worth asking for it. Existing explanations for this phenomenon attribute it to a mistaken cost computation by the help seeker, who will emphasize the instrumental cost of “saying yes”, ignoring that the potential helper also needs to take into account the social cost of saying “no”. And the truth is that, especially in face-to-face interactions, the discomfort caused by refusing to help can be very high. In short, help seekers tend to fail to realize that it might be more costly to refuse to comply with a help request rather than accepting.

A similar effect has been observed when estimating trustworthiness of people. Fetchenhauer and Dunning (2010) showed that people also tend to underestimate it. This bias is reduced when, instead of asymmetric feedback (getting feedback only when deciding to trust the other person), symmetric feedback (always given) was provided. This cause could as well be applicable to help seeking as people only receive feedback when they actually make their request but not otherwise.

Fazio, Shook, and Eiser (2004) studied something that could be reinforcing these outcomes: Learning asymmetries. By means of a computer game called BeanFest, they showed that people learn better about negatively valenced objects (beans in this case) than about positively valenced ones. This learning asymmetry stemmed from “information gain being contingent on approach behavior” (p. 293), which could be identified with what Fetchenhauer and Dunning mention as ‘asymmetric feedback’, and hence also with help requests. Fazio et al. also found a generalization asymmetry in favor of negative attitudes versus positive ones. They attributed it to a negativity bias that “weights resemblance to a known negative more heavily than resemblance to a positive” (p. 300). Applied to help seeking scenarios, this would mean that when facing an unknown situation, people would tend to generalize and infer that is more likely that they get a negative rather than a positive outcome from it, so, along with what it was said before, people will be more inclined to think that they will get a “no” when requesting help.

Denrell and Le Mens (2011) present a different perspective when trying to explain judgment biases in general. They deviate from the *classical inappropriate information processing* (depicted among other by Fiske & Taylor, 2007, and Tversky & Kahneman, 1974) and explain this in terms of ‘adaptive sampling’. Adaptive sampling is a sampling mechanism in which the selection of sample items is conditioned by the values of the variable of interest previously observed (Thompson, 2011). Sampling adaptively allows individuals to safeguard themselves from experiences they went through once and turned out to lay negative outcomes. However, it also prevents them from giving a second chance to those experiences to get an updated outcome that could maybe turn into a positive one, a more positive one, or just one that regresses to the mean, whatever direction that implies. That, as Denrell and Le Mens (2011) explained, makes sense: If you go to a restaurant, and you did not like the food, you do not

choose that restaurant again. This is what we think could be happening when asking for help: When we get a “no”, we stop asking.

And here, we want to provide a complementary explanation for the underestimation of the probability that others comply with our direct help requests based on adaptive sampling. First, we will develop and explain a model that represents the theory. Later on, we will test it empirically by means of experiments, and will elaborate on the analysis of its results.

THE MODEL

To illustrate how people underestimate the probability that others comply with their help requests, we have developed a model. Individual's choice of asking for a specific help request depends on the attractiveness of that request for them, which gets updated by the outcomes they have had when requesting it previously. Consistently with the idea of adaptive sampling, we assume that individuals will stop asking for help when they get negative responses, as the attractiveness of the request will drop, and this is what will lead to their biased estimation.

Before testing this model empirically, we examined it by means of a quite inexpensive but effective tool: Computer simulations. Simulating allowed us to find out and analyze what would happen if the scenario we set up was repeated for hundreds of runs and hundreds of trials per run. This data might have been impossible to obtain otherwise, as replicating such high number of repetitions with real participants is really expensive and time-consuming; the same holds for learning about very little variations from the initial situation.

This tool is also very helpful to detect design flaws before running the actual experiment. As the model intends to represent what the researcher has theorized about, simulation results that differ from what was expected signal that the model is wrong. The source of this deviation can be that the model has not been translated correctly into numbers, or that the initial assumptions are wrong, and therefore it is necessary to revisit them.

This model includes a variable capturing the attractiveness of each task a_i , where i can take as many values as tasks the participant can choose from. People would set the initial attractiveness values presumably according to their prior experience. As we have previously mentioned, people often miscompute the relevant costs so it is likely that the initial attractiveness of a task does not correspond to how easy it is to get help for it. In the model, we considered that these initial values would vary depending on the nature of each task and set them arbitrarily; the attractiveness of a task at time t depends on its attractiveness at $t-1$, and also on the answer to the request (if any) in the previous period, altered by a parameter b (see below). These values influence the probabilities choosing each of the tasks in the following period.

The probability $P(i)$ that the individual chooses task n for his request comes also from Denrell's (2005) work; it is an exponential version of the Luce choice model.

$$P(i) = \frac{e^{sa_i}}{\sum_{i=0}^N e^{sa_i}}$$

This function has been supported empirically as a choice model by Guadagni and Little (1983) and Yechiam and Busemeyer (2005) (both quoted by Denrell, 2005) and is often used in models of adaptive learning (for instance by Camerer & Ho (1999), quoted by Denrell (2005)).

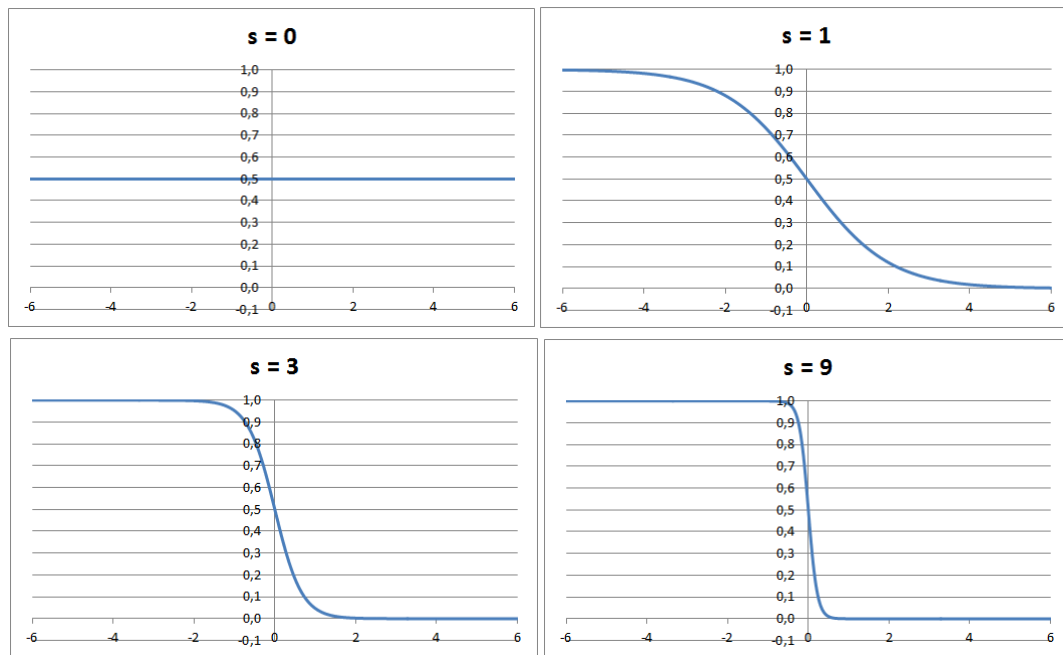
There are also two parameters that influence individual's decisions in a less obvious way. The first one is *sensitivity*, s at the spreadsheet. Sensitivity is, roughly speaking, what determines whether a tiny difference in attractiveness between two options is going to make subjects to choose it or will leave them indifferent.

It is easier to define what s is exactly by means of an example. We will suppose, for the sake of simplicity, that there are only two tasks to choose from. The probability of choosing request A is defined by the following function $P(A) = \frac{e^{s \cdot a_A}}{e^{s \cdot a_A} + e^{s \cdot a_B}}$, if we simplify this expression by dividing both the numerator and the denominator over $e^{s \cdot a_A}$, we get $P(A) = 1 / (1 + \frac{e^{s \cdot a_B}}{e^{s \cdot a_A}})$ which is the same as $P(A) = \frac{1}{1 + e^{s(a_B - a_A)}}$. From the latter equation we can see that the probability of choosing alternative A $P(A)$ depends ultimately on the difference between a_B and a_A , $(a_B - a_A)$, and the sensitivity parameter s . You can find the graphic representation of this formula in the charts below, where $P(A)$ is plotted in the vertical axis and $(a_B - a_A)$, in the horizontal one.

If s is 0, meaning that the subject is not sensitive to the attractiveness of the different options at all, alternative A will be chosen over alternative B 50% of the times, no matter the difference of attractiveness. As s increases we can see how the curve gets steeper (below). This implies that a narrower gap between the options' attractiveness is needed to make the subject decide in favor of the most attractive one.

When s tends to infinity, we will be facing an extremely sensitive subject and any difference in attractiveness, no matter how little it is, will influence the probability of choosing one alternative over the other one enormously. In this case, when alternative B is more attractive than A, the denominator will tend to infinity, and $P(A)$ will be 0. Whereas when alternative A is more attractive than alternative B, the denominator will be $(1 + e^{-\infty})$, which is 1, so $P(A)$ will be 1.

You can see the representation of how the probability of choosing alternative A $P(A)$, in the vertical axis (that goes from 0 to 1), depicted with the blue line, changes as the differences of attractiveness between the options $a_B - a_A$, in the horizontal axis (that ranges from -6 to 6), and s change.



When trying to emulate reality through our simulation, we will set $s=3$. This sensitivity variable comes from Denrell's (2005) study and he chose that value as the one that provided the best fit, understood as the minimum mean absolute deviation between the predictions of the model and the data. Low s values represent people who do not care too much about details and probably are optimistic, in the sense that hope that a bad outcome will be followed by a good one, and so that the difference in options should not be given excessive importance.

The second parameter, b , represents how much weight individuals give to the outcome of the last trial versus the previous outcomes when incorporating it to the computation of the new attractiveness of a task. A high b implies that the subject gives a lot of importance to the last outcome; namely, if he gets a negative answer when requesting task 1 in his last trial, attractiveness of 1 gets reduced by a lot, and hence it is very likely that in the next trial the individual moves from task 1 to another one even if in previous trials he got positive answers. This will be the case of impulsive people; that is people that do not take the time to reflect on what has happened but take into account the information that is more readily available, such as the outcomes of the last periods only.

The number of positive answers participants have to get in order to complete the experiment was set under the label of "stop". This number came from partially mirroring Flynn & Lake's (2008) experiment. Using this design makes that the number of positive answers divided by the total amount of requests is a biased estimator of the probability of compliance. Putting it differently, assume that instead of obtaining 3 "yes", the participant is told to make 10 requests. This participant happened to get 3 positive answers anyway. When computing the probability of success, we will keep 3 as the numerator, but the number at the denominator will increase, as we will be taking into account all the trials that came after the third "yes" and that we would be ignoring in the previous setup. The probability of compliance computed first will be higher than the second one. This is not a problem as long as the experimenters acknowledge it and make their conclusions accordingly.

In each period t , participants will choose the task n that is most attractive, with probability $P(i)$. Once they have asked, they will get an answer to their request that comes given by the real probability of success $rP(i=y)$. If the answer is positive, the attractiveness of that task will be upgraded, and it is likely that the participant chooses it again; otherwise, it will be downgraded, and it is likely that the participant stops choosing that one, namely he switches to another task. We simulated this setup for large numbers of runs and trials and then we obtained the simulated experienced probability of successful request for every task i , $sP(i=y)$.

Once we have finished with the description of the model components, the next step is to change the initial setup to simulate different scenarios, observe how the different probabilities (especially $sP(i=y)$) change, and then make predictions.

For our experiment, we will count with two different conditions whose outcomes will be compared. Participants in Condition 1 will be forced to screen, as they will only have one task available. As information about real probabilities is contingent to approach behavior, they will get a larger sample from where they will be more likely to estimate a probability which is closer to the real one than the ones in the other condition. Participants in Condition 2 will be allowed to stop screening one task by means of switching to another task, that is, to sample adaptively according to the experiences they get at prior first trials. What we expect to see in the simulations is this difference in the experienced probabilities for one task of the two groups. We presume that the experienced probability of success for those in Condition 2, where participants can stop sampling one task, will be lower than for those in Condition 1, where they are forced to screen. In order to set up these conditions in the excel spread-sheet, we needed the real probabilities of success for all the tasks; it is very difficult for the experimenters to know the 'real' probabilities of success for the tasks, so we approximated them according to the samples we have.

The samples obtained by participants in these two conditions will also differ by the last observation included. In the first condition, participants will end their sampling process with a "yes", as they will complete the experiment once they get three positive responses. The higher amount of information is likely to make their experienced probability closer to the real one. But the fact that the last observation they obtain laid a positive outcome may make, in samples in which the average size is 6.8 observations, this probability is overstated. Participants in the second condition, however, are likely to stop screening one task to switch to another one when they get a "no", and this, together with their lower sample size, will make that they experience and report lower probabilities.

We have not included in the simulations the estimates of the probabilities of success subjects will report after finishing the experiment. However, we can expect them to be lower than the experienced ones, due to the learning asymmetry towards negative valenced objects (explained by Fazio, Shook, and Eiser (2004)). According to them, people will tend to remember more the negative events (as getting a "no" as response) and therefore they will underestimate even further their probability of success.

We changed the setup of the simulations in order to see what would happen in different scenarios. We obtained every simulated result by running 10 runs of simulations with 500 trials per run. In the first condition participants have to sample no matter how sensitive they are to the last outcome, or how fast they update their beliefs, or how attractive they perceived the task to be initially, so most of the times altering this background will not make sense. We will alter the values of these parameters and variables in the second condition. We will display our results in charts or graphs, where you will find the values of the parameters 's', 'b', and 'stop', also the attractiveness 'a', the simulated probability of success 'sP(i=y)' and the real one 'rP(i=y)' and the simulated probability of asking for the task 'sP(i)'.

Simulation 1: Realistic setup

In the first simulation, we wanted to emulate a realistic setup that could resemble the experiments we were to run. In this setup, the probabilities that participants get a positive or a negative answer are realistic, we obtained them at the pre-testing phase of the experiment (Pilot Experiment 2, see on page 16): They are 43% for the cellphone task, 57% for the picture task, and 31% for the escort task. The value of s is 3, which we know is the one that fits reality better; the value of b is 0.6, implying that people give some more weight to our last observation but not that much more, and participants, as in the experiments, will have to stop after getting 3 "yes". We assumed that the three tasks were equally attractive; this is probably not the case for an individual, but we would achieve that in a large group thanks to random assignment, as we would get people with different preferences, which would cancel each other out and even out when taking the average. We expect that our predictions about the probabilities that you read just above are met.

For the cellphone task, which is the 'control' task (the one we will compare across conditions), the simulated probability of success in Condition 1 (0.501) is higher than the one in Condition 2 (0.409). This supports our initial prediction. It is strange, though, that this probability in Condition 1 is higher than the real one, 0.430; the underlying cause of this could be what we have explained before about the last observations included being successes. This would not be affecting the results of the second condition because there, as participants can switch to sample another task, the sample does not need to be biased towards successes.

Simulation 2: Neutral setup

In this simulation we replicate what would happen if participants faced a situation we have called 'neutral'. In this situation, we keep $s=3$ as it provides a good fit with reality, but set b at its middle value, 0.5, implying that subjects give the same importance to their last observation than to the previous one. We make the three alternatives equally attractive, and make that they all have the same probability of success of 0.5.

CONDITION 2 – TREATMENT				
	A	rP(i=y)	sP(i=y)	sP(i)
Phone	0.5	0.500	0.455	0.341
Picture		0.500	0.431	0.331
Escort		0.500	0.447	0.333

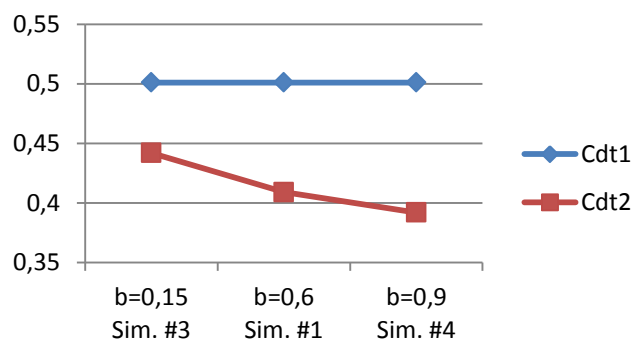
	CONDITION 1 – CONTROL	
	rP(i=y)	sP(i=y)
Phone	0.500	0.588

This simulation offers very good evidence that what we are predicting holds, as the probability in Condition 2 (0.455) is lower than the one in Condition 1 (0.588).

In this case, and in order to give stronger support to what we have found, we have included the simulated probabilities for all the tasks. When we rank the tasks according to how often they have been sampled, we observe that this same ranking is valid for how close the experienced reality is to the real one; so the more you sample a task, the more likely you are to experience the real probability of success. Namely, we see that in a context where none of the options is discriminated in favor or a more attractive one, participants will experience a success probability closer to the real one.

Simulations 3 and 4: Varying parameter b

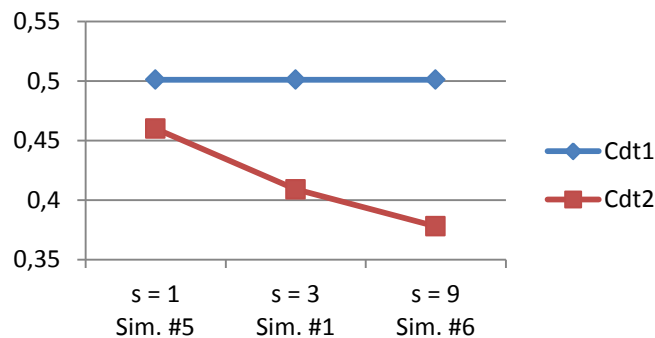
In this simulation, we are going to use the same probabilities of success than in the realistic one, and we are going to alter parameter b , to make subjects put mostly all the weight on their last observation (high b) or almost nothing but rather consider the previous ones (low b). In the horizontal axis of the graph we have placed b , and in the vertical one, the simulated probabilities of success. Each of the two lines plots the probability for a given value of b for each condition.



Once again, we see our predictions taking place in the simulations: The probabilities from Condition 2 are lower than the one from Condition 1, no matter which value b takes. Also, we observe the overestimation with respect to the real probability in the single-task condition. When focusing on Condition 2, we see that the larger b gets, the further the experienced probability deviates from the one in Condition 1, that is, the higher the underestimation. When $b=0.15$, the underestimation gets reduced, actually turning into a little overestimation (0.442). This is caused by the very little weight that is given to the last trial, which will make that even when it laid a negative outcome, attractiveness will not change too much, so the individual will continue sampling that task, getting closer to a forced-sampling scenario. Going along with that, we can see that the moment when the simulated experienced success probability (0.392) is further to the real one (0.43) and the one at Condition 1 (0.501) is when $b=0.9$. Hence, subjects who mainly about the present and have little memory will be the ones suffering the most from this underestimation.

Simulations 5 and 6: Varying sensitivity (parameter s)

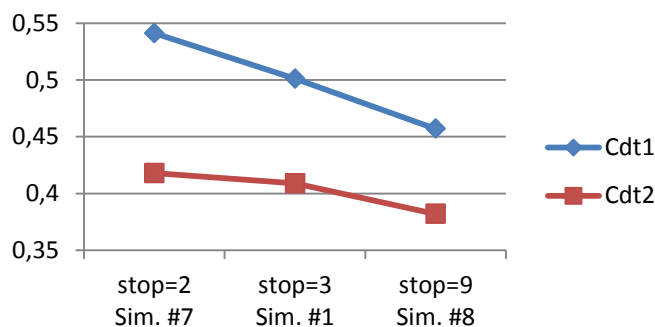
In this simulation, we are showing what people with really low and really high sensitivity would experience. We have placed s on the horizontal axis, and the rest is the same as in the previous graph.



The effect of this change in s is very similar to the one achieved with variations in b ; this, hence, suggests that these parameters change effects are related. It is very interesting to observe that when individuals are not very sensitive ($s=1$), as they tend to stay sampling the same option even when they get rejected (same as when b is low), the experienced probability of success (0.460) gets closer to the real one (0.430) and even exceeds it; low sensitivities are therefore be mimicking forced sampling scenarios. In the case of high sensitivity ($s=9$), we observe the opposite; as individuals stop sampling one task as soon as they get a “no”, it is rarer that they stay long enough sampling one task to get an accurate estimate of its probability of success.

Simulations 7 and 8: Varying target (“stop”)

In this case, what we are varying is the target which is parameter “stop” in the simulations. The target will take the values of 2, 3, and 9.



We chose not to make it take the value 1 so that the experiment was not distorted, as with such a short screening period, there would be no learning time and the probability predictor would be biased because of the little trials. It actually seems that this is what is happening when the target is 2 successes: The overestimation in the control condition gets sharper than in the realistic setup, following the logic mentioned for $stop=1$. Continuing with the analysis of Condition 1, we see that the higher the target, the more sampling opportunities subjects have, and so the closer they get to the real probability.

The results we obtain from simulating Condition 2 are not exactly what we were expecting. While our prediction about the comparison of the probabilities between conditions still holds, we see that the more screening opportunities we give to subjects, the lower the experienced probability of success will be. This allows that the ‘underestimation gap’ between the conditions remains approximately the same, but it is definitely a phenomenon that is difficult to explain.

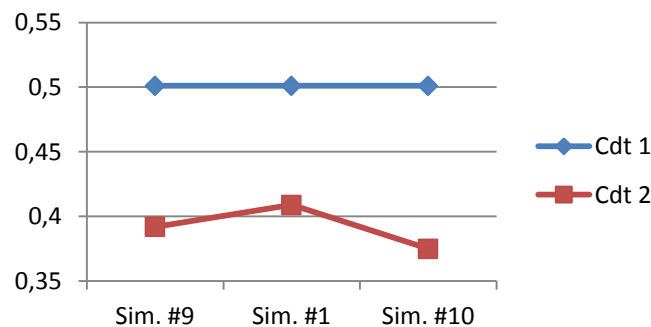
Simulations 9 and 10: Varying relative attractiveness

In the following simulations, we are going to change the relative attractiveness of the options to make it go in accordance to the real probabilities of success or in the opposite direction. These changes may have important implications for our decision of which tasks to include in our experiment. In simulation 9, we wanted attractiveness to mimic the probabilities (as if participants actually knew about them), while still giving a proportionally higher attractiveness to the most ‘successful’ task. In simulation 10, we wanted to put this task as the least attractive, while leaving the individual indifferent between the other two. In Condition 1, experienced probabilities will not change according to attractiveness, as it is not included in the model because subjects in this condition are forced to sample and cannot choose among several options, so the simulated probability of success will not deviate from the one at the realistic setup.

		Simulation #9			Simulation #1			Simulation #10		
		CONDITION 2 – TREATMENT								
		rP(yes)	A	sP(i=y)	sP(i)	A	sP(i=y)	sP(i)	A	sP(i=y)
Phn	0.430	0.3	0.392	0.303	0.5	0.409	0.310	0.5	0.375	0.311
Pic	0.570	0.6	0.561	0.466	0.5	0.545	0.444	0.2	0.569	0.441
Esc	0.310	0.2	0.259	0.266	0.5	0.277	0.230	0.5	0.236	0.231

		CONDITION 1 – CONTROL					
	rP(yes)	A	sP(i=y)	A	sP(i=y)	A	sP(i=y)
Phn	0.430	X	0.501	X	0.501	x	0.501

Even when that is not our focus, we have considered interesting to show the outcome for all the tasks, as it is reinforcing the idea that the more screening, the lower the underestimation with respect to the real one. We also observe that our prediction of an underestimation of the success probability of the cellphone task in Condition 2 with respect to Condition 1 takes places in all cases, no matter what the relative attractiveness are.



Moving onto the comparison of the different scenarios simulated, we see that when attractiveness scores are not uniform, the underestimation is bigger, no matter whether they are aligned or not with real probabilities. In the case that attractiveness reflects the true probability, still making relatively more appealing the most successful one, the experienced probability of success gets further down than the real one than in the realistic setup, making wider the underestimation. This makes sense, as the higher initial attractiveness pushes individuals to focus on the successful task, leaving the other unattended in comparison, and so, there is not enough opportunity to sample the control task. We observe the same trend at Simulation 10, but much stronger; this result is shocking and it is difficult to figure out why the opposite situation is causing the same effect, but even more intense.

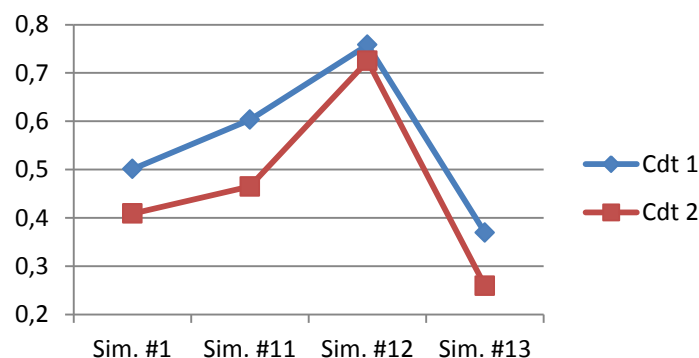
Simulation 11 to 15. Varying tasks difficulty (probabilities of success $rP(i)$)

In the last simulations, the parameters are $s=3$, $b=0.6$, and $stop=3$, as in the realistic setup, once again. In this case we are going to be manipulating the 'real' probabilities of our tasks. This will help us to get a feel of whether different combination of tasks would lay different results.

The first set of scenarios is comprised from simulations from 11 to 13, which will also be compared to the outcome of the realistic setup in simulation 1. In Simulation 11, the success probabilities for all the tasks will be the same and neutral (50%). In Simulation 12, the most successful task will be the control one, which will be more than twice as successful as the other two. In simulation 13, we will emulate the same scenario, but the overly-successful task will not be the control one but one of the other two.

		Simulation #11			Simulation #12			Simulation #13		
		CONDITION 2 – TREATMENT								
	A	rP(i=y)	sP(i=y)	sP(i)	rP(i=y)	sP(i=y)	sP(i)	rP(i=y)	sP(i=y)	sP(i)
Phn	0.5	0.500	0.465	0.341	0.700	0.725	0.610	0.300	0.259	0.250
Pic		0.500	0.436	0.327	0.300	0.241	0.199	0.700	0.713	0.615
Esc		0.500	0.486	0.344	0.250	0.194	0.184	0.250	0.202	0.181

		CONDITION 1 – CONTROL					
	A	rP(i=y)	sP(i=y)	rP(i=y)	sP(i=y)	rP(i=y)	sP(i=y)
Phn	x	0.500	0.603	0.700	0.758	0.300	0.369



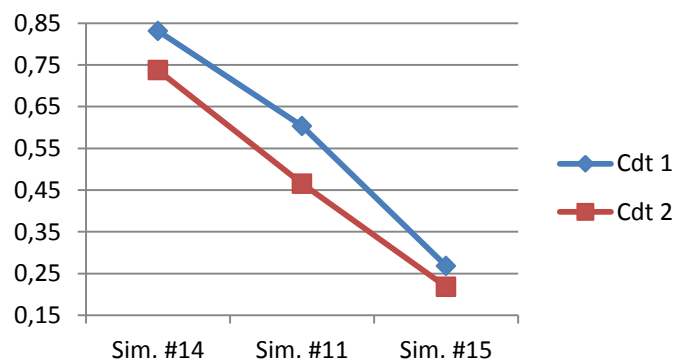
The outcome of Simulation 11 is very similar to the one of the neutral scenario, (the only difference between them is b , which differs only by 0.1); in Condition 2, all the tasks are approximately equally sampled, the probability gets close to the real one, and it still is under the one experienced when sampling was forced, as we predicted. The divergence in the experienced probability and the real ones must come from a lack of learning time, so this probability should tend to 0.500 with time.

Simulations 12 and 13 enclose the same success probabilities, but have them assigned to different tasks. In the first one, the control task is the most successful one; in the second one, the most successful task is the second task, and the control one is the next one in terms of success. We see that the outcome varies in accordance to the change in the real probability (see the peak), but the most important here is to acknowledge how the simulated probabilities of success vary depending on which task has been assigned the outstanding real probability of success. When the control task has a 70% success rate, and the next one a 30%, the simulated probability of success for the latter is 24.1%; whereas when the one with the 30% success rate is the control task, the first one that is screened, the simulated success rate goes up to 25.9%. Tasks with identical success probabilities will lead to different experienced probabilities just because of the sampling order; more specifically, we see that the first task, no matter whether has a 30 or a 70% of success rate, will get a higher experienced success when it is first (control) than when is the second task. This should be taken into account when designing our experiments, and it will definitely be important when we analyze the results we get.

In the second and last set of simulations (14 and 15 vs. 11), we will compare the scenarios where the probabilities of success are the same for all the tasks, but are either high (80%), neutral (50%), or low (20%).

		Simulation #14			Simulation #11			Simulation #15		
		CONDITION 2 – TREATMENT								
	A	rP(i=y)	sP(i=y)	sP(i)	rP(i=y)	sP(i=y)	sP(i)	rP(i=y)	sP(i=y)	sP(i)
Phn	0.5	0.800	0.737	0.326	0.500	0.465	0.341	0.200	0.218	0.332
Pic		0.800	0.798	0.334	0.500	0.436	0.327	0.200	0.206	0.330
Esc		0.800	0.764	0.342	0.500	0.486	0.344	0.200	0.221	0.335

		CONDITION 1 – CONTROL					
	A	rP(i=y)	sP(i=y)	rP(i=y)	sP(i=y)	rP(i=y)	sP(i=y)
Phn	x	0.800	0.831	0.500	0.603	0.200	0.268



We can still observe how our predictions, no matter at which level probabilities are, still hold. In these cases, we see that the proportion of sampling time is evenly split among all the tasks. Moreover, in the first two simulations, when the probabilities are either high or neutral, the experienced probabilities in Condition 2 are close to the real ones, and still under the one from Condition 1, which is, as usual, an overestimation.

It is very curious what we have found that would happen in the case all the probabilities were equal and low (20%). The experienced probability of success in Condition 2 would already be an overestimation, and it would be closer than ever to the one in Condition 1, which surpasses the real one by almost a 7%.

The purpose of running these simulations was to approximate what we could expect in a real life setup. From the results we obtained, we can extract the following results that we intend to validate by means of the experiment.

Result 1: *The experienced probability of success from subjects in Condition 2 will be lower than the one from subjects in Condition 1. This holds for all values of b , s , “stop”, attractiveness, and real probabilities of success.*

Result 2: *In a context where none of the options is discriminated in favor of a more attractive one, participants will experience a success probability that is closer to the real one.*

Result 3: *The higher b , the higher the underestimation of the success probability in Condition 2 with respect to Condition 1. The same happens with parameter s , and “stop”.*

Result 4: *No matter the real success probability, the first task to be screened will always get a higher experienced success probability than it would get if it was not the first one to be screened.*

Result 5: *When the chances of screening are very high, such as in the forced sampling condition, or when one task is much better than the other ones, there can be an overestimation of the probability of success. When this happens, the probability in Condition 2 will still be lower than the one from Condition 1.*

Proof of these results will come (or not) with the experiment and posterior data analysis.

THE EXPERIMENT

The experiment is what helps us to test whether the theoretical model holds empirically or not. Even when you have spent a lot of time and effort in planning the smallest details of the experiment design, you may need to change some parts of it, as it is impossible to predict how exactly participants are going to react to incentives, whether they will interpret the instructions the way you wanted, etc.

As we have mentioned before, the contribution that this experiment is testing is whether adaptive sampling can be an explanation of people's underestimation of how much others are going to comply with their direct requests for help.

Our experiment partially replicates the one that Flynn and Lake did in 2008. As a little reminder, they found out that people did underestimate the probability that others would comply with their direct requests for help and explained it in terms of social and instrumental costs (Flynn & Lake, 2008). What we want to do is to show that this underestimation is also caused by the adaptive sampling mechanism. We have adapted their experiment by adding up different tasks in the same experiment and letting participants use adaptive sampling, i.e., switching from one to another as a way of quitting the first one, in which they feel they will not get help. After that, we will ask participants to report the probability they had to get help for the task they have abandoned and we hope to find that it is indeed lower than the one they participants report when they cannot use adaptive sampling.

We will go through the three milestones of experimentation that Fisher (1935, 1956) (quoted at Thye (2007)) established in order to explain our experiment in some more detail. The first one is the *random assignment* of participants within the two conditions, and it is used to avoid that irregularities in participants are only present in one of the groups. Our assignment was not completely random but it would still work to hold constant spurious cases; we let participants sign up for the session they preferred and then we assigned a condition to every session. This also helps that participants do not know about the existence of two different conditions.

The second one consists on introducing a *manipulation*. This manipulation is measured as the *independent variable* of the study, and it differs between conditions; in this case is the possibility of stopping sampling. We already talked about the two conditions of this experiment before, but we will provide a little reminder here tied to the theory basis. Our first condition is the one of the *control group*, in which subjects are forced to screen, that is, to request compliance for the 'control' task. In the simulation for this condition there was no attractiveness measure but one contemplating the "psychological effect" of the answer to the request (1 for "yes", -1 for "no"). The second condition is the one of the so-called *treatment group*, in which participants can choose to stop requesting a specific task and switch to another one. This is a representation of the situation when people decide to stop asking (or sampling) after getting some negative answers, as they are failing to assess correctly the probability of getting a positive answer.

Finally, the last essential component of research is about having a *controlled measurement*; that is, measuring the dependent variable in a controlled environment. Our *dependent variable* is the experienced probability of getting help when requesting the 'control' task. What we want to prove with this experiment is that the independent variable makes that the dependent variable for Condition 1 is higher than the one for Condition 2.

The tasks that were used for the pre-testing sessions were intended to be the most similar possible to the ones at Flynn and Lake's (2008) experiment: Questionnaire (ask people to fill in

a 5-10 minutes long questionnaire), Escort (ask people to walk us to the nearest metro station), and Cellphone (ask people to let us borrow their cellphones to make a short phone call). This is because we assumed that they gave extensive thought to their design, and we therefore could benefit from it. However, we did not blindly trust their tasks and also to check if the same conditions could be replayed, we did some pretesting and ran pilot experiments.

Pretest 1

Procedure: The first pretesting consisted on going next to beach in Barcelona, the place where the real experiment was going to be run, and try the tasks the way they were at Flynn and Lake's paper. The participants were an experimenter and a confederate, 21 year old university students, female and male. We followed the script that participants of Flynn and Lake's experiment had, which directly asked for the request to the potential helper.

Results: From the people we approached to ask them to fill in a questionnaire, 30% agreed to do it. 20% of the cellphone tasks were answered positively. And regarding the escort task, it would work as long as you approached someone who was on their way to the metro.

Discussion: Getting a questionnaire filled was harder than it looked but a 30% success rate was what we were looking for. Borrowing a cellphone was considerably harder; we assume that it is very weird for someone to get approached by a stranger that directly asks you for your phone and does not even say a reason, it turned weird even for the requester. This gets even harder when you are in Barcelona, where pickpocketing is really common, and potential helpers were suspicious. More observations, gathered by a higher number of participants so as to increase the random assignment, should be gathered, so a second pretesting session was done.

Pretest 2

Procedure: Three 20 year old university students, two males and one female participated in this pretest session. In this case, the cellphone task was removed and instead, the delivery task was introduced, in which participants would have to ask someone to take a paper to the experimenter, who would be sitting a few meters away.

Results: The delivery task achieved a compliance rate of 36%. We combined the results of both pretest sessions and the questionnaire task reached 38% of compliance and the proportion of positive answers to the escort request was 33%.

Discussion: The results from this pretest were good because the tasks had similar success rates and that would avoid that some participants focused on "the easy one" and did not switch.

Pilot Experiment 1

Procedure: The first pilot experiment was run with 10 participants, 8 females and 2 males, 18-19 years old, all of them students of International Business Economics that were compensated with some participation grade bonus. The first 5 participants were assigned to the treatment condition, and the other 5 ones to the control one. Participants had to do a trial of each task first, and then they could choose what to ask for; this was like this so that even when the

participant considered the control task as the hardest, would try it at least once. Participants, based on their experience during the experiment, reported an estimated probability, by answering to an open question about the number of people they would need to approach to get three people to comply with the escort task. The control condition session had to be cancelled after a while because it started raining.

Results: The probabilities of success for each task when participants were in the treatment condition were 92% for the questionnaire, 50% for the delivery, and 14% for the escort. The 'control task', the one whose success rate we would compare between groups, was the escort one. The probability of success for this task in the control condition was 21%, higher than in the treatment condition. Participants' average success probability estimates were 18% in the control condition and 19% in the treatment one.

Discussion: The fact that the experienced probability of getting a "yes" is lower in the 'three tasks condition' (treatment condition) than in the 'one task condition' (control condition) is in line with our predictions. In the 3 tasks condition they experienced a lower probability of success, so the estimates should vary according to that, but surprisingly they do not; the average probability estimate is higher in the treatment condition. However, we cannot know if this is just a coincidence because the sample size is not big enough, and because the observations between the two groups might differ as the ones in the control group could not finish the experiment. In addition, we observed that the values of their estimates were only 'round' numbers (10, 15, 20, and 25) with the exception of 12. We suspect that participants tended to round their estimates (for instance, they would tend to answer 20 instead of 19 or 21) due to the open-ended format of the question. This could also be biasing their reports. We saw, therefore, the need for changing the answer format, and we moved onto a chart with numbers ranging from 3 to 100+, so that they could pick 20 as easily as 19 or 21.

The questionnaire and delivery tasks were too easy in comparison to the escort, which made participants to mostly make their help requests about them. As this would not have let us see the effect on the main task (escort), because they might be capturing too much attention, we changed them.

Pilot Experiment 2

Procedure: The second pilot experiment was run with 6 participants, 5 females and 1 male, from the same characteristics than the ones of the first pilot. All of them were in the treatment condition, as we needed to test the new tasks. In this case, the escort task was maintained but the other two were substituted by a version of the original cellphone task, and a new one that we call "picture task". In the cellphone one, participants asked if the potential helper would be willing to do them a favor, and then asked them to text the experimenter on their behalf. The way of making the request was 'less aggressive' than in the first version of the task, and it did not involve the helper actually giving their phone to the requester. The picture task consisted in getting potential helpers to get a moustache drawn in their face by the help seeker and get a picture taken. Participants told potential helpers that they would delete the picture right after and that they could remove the moustache with wet wipes. Again, their three first trials were one for each task, and then they could ask freely.

Results: The experiences probabilities of success were 31% for the escort task, 57% for the picture task, and 43% for the cellphone task. The average probability of success for the escort task participants estimated was 11%. With the new format of the post-experiment questionnaire (see at Annex, p. 25), we observed more diverse estimates such as 4, 14, or 86.

Discussion: The new tasks were easy to monitor (we would see the participant coming with the escort and the picture of the person with the moustache, and receive the text message). They also laid good success rates that were close enough from each other. As the escort success rate was a bit lower than the other two, we decided to make the cellphone task the control one in order to make sure that the control task would not be the one which potentially could be left untested.

Pre-experiment survey

Procedure: At this point, we had decided that the control task would be the cellphone one and we had tested that the way in which we ask participants to estimate probabilities would not give us biased answers, so we distributed a questionnaire very similar to the post-experiment questionnaire. We wanted to get estimates from people who have had no experience in that specific task. At the same time, we thought that asking subjects about this before they tried the assignment could influence their behavior during the experiment, so that is why we only approached people who had not or would not take part in the experiment. This questionnaire differed from the post-experiment one in that it included the instructions for the help request.

Results: We got 57 questionnaires back. 77% of them were responded by females, and the rest by males. Our respondent base was constituted by students that were on average 21, 86% of whom were Spanish. The average number of people they estimated they would need to get three “yes” for the cellphone task was 20.67, that is, they estimated a probability of success of 14.5%.

Discussion: It is interesting to compare the estimates of subjects who had not gone through the assignment (these ones), with the ones who did (respondents of the post-experiment survey) to see whether this experience had changed their perceptions. We distributed this questionnaire right after changing the control task from being the escort to the cellphone. That is why we could not compare the results obtained from these questionnaires against any data until we ran Experiment 1. Discussion about the estimates will therefore come at the Discussion section of the next experiment.

Experiment 1

Procedure: For the final experiment, we recruited 56 participants, of which 30 never showed up, 7 decided not to participate after reading the instructions (see at Annex, p. 22-23), and 2 did not finished the assignment. Those who took part in the experiment were students, 21 years old on average, 13 females and 14 males, evenly distributed between the two conditions. Participants who completed the assignment were rewarded by having the chance to win one of three 50€ amazon gift cards. The meeting point of the experimenter and the participants was a café placed next to the beach promenade, where the experiment was run.

We did not see the participants while they were asking the favors, but we monitored compliance with the tasks in the following way: For the phone task, we received the texts and checked whether the ones received from each participant coincided with the ones they reported; for the picture task, we asked participants to show them to us; and for the escort task, we required them to wait for our wave, once we saw them with their escort. We chose the cellphone task to be the control task, the one whose probability we would compare in the two conditions, because we found it very easy to monitor.

Results: The results can be seen below in the table displayed by SPSS. The average experienced success probability (P_success) for the phone task in Condition 1 was 0.280, and the one in Condition 2 was 0.650. The success probabilities they estimated were 0.188, and 0.298 respectively.

Group Statistics					
	Condition	N	Mean	Std. Deviation	Std. Error Mean
P_success	1	12	,279800	,1685797	,0486648
	2	15	,650000	,4799554	,1239239

(General) Discussion and Data Analysis: Before diving into the analysis of our predictions to check whether our predictions hold, I will go back to the pre-experiment questionnaire and contrast the estimates from participants who did the experiment, and those who did not. Subjects who had not gone through the ‘requesting for help’ experience, reported an estimated probability of success of 0.145, while the estimates coming from experiment participants were higher (0.188 from Condition 1, and 0.298 in Condition 2). This is consistent with our prediction that the more sampling, the better fit of the predicted probability with the real one. However, this goes in the opposite direction than that, once we consider that in Condition 1 participants were forced to sample, whereas in 2 they were not, but the ones in 2 still got higher estimates of success than in 1.

Then, in order to decide whether we can give credit to this data or not, we should see whether the difference between groups is statistically significant or not. In order to see whether there are significant differences in the means of experienced probability of success between the two unrelated groups of participants (Condition 1 and 2) we can use the *independent samples t-test* (Sekaran & Bougie, 2010). We will first assume that our data follows a normal distribution. Then we will see, by means of a *Levene’s Test* that SPSS will run for us, if we can assume that there are equal variances or not.

Levene's Test for Equality of Variances		
	F	Sig.
P_success Equal variances assumed	30,269	,000
Equal variances not assumed		

As the p-value we have found (0.000) in the *Levene’s Test* is lower than the confidence level we set (0.05), we can reject the Null Hypothesis assuming equal variances. Because of this, we

will continue the analysis looking at the second line of the output of the t-test of equality of means.

t-test for Equality of Means

F	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
30,269	-2,541	25	,018	-,3702000	,1456903	-,6702548	-,0701452
	-2,781	18,103	,012	-,3702000	,1331368	-,6497960	-,0906040

The p-value obtained from the *t-test* (0.012) is lower than 0.05, meaning that we can reject the Null Hypothesis of equality of means, and state that the difference we have found between the average probabilities of success of the treatment group and the control group is statistically significant, so we can move onto the analysis.

These results go in the opposite direction of our predictions: Contrary to what we had predicted on the basis of the theory, our simulations, and our trial experiments, the experienced probability of success for subjects in Condition 2 (on average), is significantly higher than the one from Condition 1. However, we would like to cast some doubts on their validity because of the following reasons. Due to the high absenteeism among the participants that signed up for the experiment, the sample size we obtained is not big enough and so the results could be just one part of the picture, that with a bigger sample size might go in the opposite direction; more observations are needed.

Other reasons esteeming from the experimental design could as well be biasing the results, such as curiosity that participants could feel for trying one of the tasks no matter how difficult they feel it is, how many “no” they have got, and disregarding the goal of the experiment. This meets the characteristics of the picture task in our experiment, as it might come across as a fun thing to do. It could be therefore that this task had attracted participants’ attention and therefore their request attempts, and that it had retained them afterwards, due to the high probability of success. According to our simulations, this can result into a slight overestimation of the estimated probabilities with respect to those experienced when the control task is the most successful; it remains uncertain whether this would be enough to explain an overestimation such as the one we have found. A way to overcome this situation is to exclude this task, although, according to our simulations, we could still find the effect we predicted by placing it as the control task. The findings of Simulations 12 and 13, stated in Result 4 (p. 13), might also be playing a role: The fact that the task that we are comparing is the first to be screened might make that we are observing an overestimation of the experienced probability this task would get if it was second. This might explained another bit of the high experienced probability we have observed.

Secondly, it is worth noting a pattern of responses we obtained from some participants at Condition 2 that could be biasing the results: Participants got three “yes” in the first three attempts, meaning that they probably did not sample all kinds of people and see whether they complied or not, but only made the requests to those people they thought they were most

likely to comply. The fact that they were waiting for the 'right person' to walk by is supported by the long time it took them to finish the experiment, given that they only made 3 requests. We were expecting this to happen, but not in such an extent that they did not have time to get any negative response and sample consequently afterwards. Just with the aim of seeing whether data supports our claim on this, we excluded participants with this pattern from the data and we were left with 9 participants. The experienced probability of success was substantially lower, 0.417 (vs. 0.650), while their estimated probability decreased until 0.239 (vs. 0.298).

All in all, more observations and some changes in the experiment design are still to be made in order to be able to accept or discard our predictions.

REFERENCE LIST

- Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, 112, 951–978.
- Fazio, R. H., Shook, N. J. & Eiser, J. R. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, 87, 293-311. doi: 10.1037/0022-3514.87.3.293
- Fetchenhauer, D., & Dunning, D. (2010). Why So Cynical?: Asymmetric Feedback Underlies Misguided Skepticism Regarding the Trustworthiness of Others. *Psychological Science*, 21(2), 189–193. doi:10.1177/0956797609358586
- Flynn, F. J. & Lake, V. K. B. (2008). If you need help, just ask: Underestimating compliance with direct requests for help. *Journal of Personality and Social Psychology*, 95 (1), 128-143.
- Fisher, R. A. (1935). *The design of experiments*. London, UK: Oliver and Boyd.
- Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh, UK: Oliver and Boyd.
- Hogg, M. A. & Vaughan, G. M. (2010) *Essentials of social psychology*. Essex, UK: Pearson Education Limited
- Mens, G. le. & Denrell, J. (2011). Rational learning and information sampling: On the “naivety” assumption in sampling explanations. *Psychological Review*, 118 (3), 379-392.
- Sekaran, U. & R. Bougie, (2010) *Research Methods for Business: a skill building approach* (5th ed.). Chichester, UK: Wiley
- Thompson, S. (2011) Adaptive sampling [PowerPoint slides]. Retrieved from http://www.mathstat.helsinki.fi/msm/banocoss/2011/Presentations/Thompson_web.pdf
- Thye, Shane R. (2007). Logical and Philosophical Foundations of Experimental Research in the Social Sciences, in M. Webster & J. Sell (Eds.), *Laboratory Experiments in the Social Sciences*. London, UK: Elsevier

ANNEX

Instructions – Condition 1

Dear participant,

The research team wants to thank you for taking part in this experiment. We also want to encourage you to be honest so that we can count on good quality data for our research.

In this experiment, you will have to ask people for the favor described below. Your goal is to get three “yes” in total, that is, three people helping you. When you ask for a favor, people might comply or might refuse to comply with your request. You have to keep trying until three people have complied with your requests.

Please report your experiences in the table on the last sheet of these instructions. Every time you ask for a favor, please write down whether the request was successful or not. If your target does not want to talk to you, this counts as a ‘No.’

You can ask in Spanish, Catalan, or English. This choice depends on the language you prefer and the language that your potential helper speaks. When asking for help, it is important that you do not deviate from the scripts you will find below.

Finally, you can only approach people that you DO NOT know already. That is, you should ask favors from people that are complete strangers to you.

PHONE

This favor consists in asking someone to text the experimenter (number: 670527760) on your behalf. The text message needs to include your location and (most importantly) your participant number. This task will only be successful if the experimenter has received the text message, so stay with your helpers while they write it and send it to make sure that they are really complying with your help request.

When you approach somebody, you will have to follow the following script:

SPANISH: Disculpa, ¿podrías hacerme un favor?

CATALAN: Disculpa, podries fer-me un favor?

ENGLISH: Excuse me, could you do me a favor?

If the answer is “yes”, then you have to tell them:

SP: Necesito que envíe un mensaje de texto de mi parte.

CT: Necessito que envii un missatge de text de part meva.

EN: I need that you send a text message on my behalf.

If they press you for details, you have to offer minimal information:

SP: Puede escribirlo usted mismo. Es para un estudio de psicología.

CT: El pot escriure vosté mateix. És per un estudi de psicologia.

EN: You can type it yourself. It's for a psychology study.

To begin the experiment, please, go to Experimenter 1, who is sitting at the café of the CEM Marítim (see the attached map). Once there, meet the experimenter, who will be holding a red

folder from UPF and a red backpack. The experimenter will give you some further instructions. Once you have finished, please, meet the experimenter at that spot again in order to complete a short post-experiment questionnaire and to submit the record of your help requests.

Thank you again and have fun!

The Research Team

[Page break]

Trial Number	Compliance (circle Y or N)
1	Y / N
2	Y / N

... till 30, which was the amount of trials that filled up a sheet.

Instructions – Condition 2

Dear participant,

The research team wants to thank you for taking part in this experiment. We also want to encourage you to be honest so that we can count on good quality data for our research.

In this experiment, you will have to ask people for some favors described below. Your goal is to get three “yes” in total, that is, three people helping you for any of the tasks. When you ask for a favor, people might comply or might refuse to comply with your request. You have to keep trying until three people have complied with your requests.

It is very important **that you begin with asking once each favor**. That is, in the first three trials, you need to ask once for an escort, once for somebody to get a picture with a drawn moustache, and once for their cellphone. After these first three trials, and no matter whether your requests were successful or not, you are free to ask for the favor of your choosing.

Please report your experiences in the table on the last sheet of these instructions. Every time you approach somebody in order to ask for a favor, please write down the type of favor (‘escort’, ‘picture’ or ‘phone’) and whether the request was successful or not. If your target does not want to talk to you, this counts as a ‘No.’

You can ask in Spanish, Catalan, or English. This choice depends on the language you prefer and the language that your potential helper speaks. When asking for help, it is important that you do not deviate from the scripts you will find below.

Finally, you can only approach people that you DO NOT know already. That is, you should ask favors from people that are complete strangers to you.

PHONE

This favor consists in asking someone to text the experimenter (number: 670527760) on your behalf. The text message needs to include your location and (most importantly) your participant number. This task will only be successful if the experimenter has received the text message, so stay with your helpers while they write it and send it to make sure that they are really complying with your help request.

When you approach somebody, you will have to follow the following script:

SPANISH: Disculpa, ¿podrías hacerme un favor?

CATALAN: Disculpa, podries fer-me un favor?

ENGLISH: Excuse me, could you do me a favor?

If the answer is “yes”, then you have to tell them:

SP: Necesito que envíe un mensaje de texto de mi parte.

CT: Necessito que envii un missatge de text de part meva.

EN: I need that you send a text message on my behalf.

If they press you for details, you have to offer minimal information:

SP: Puede escribirlo usted mismo. Es para un estudio de psicología.

CT: El pot escriure vosté mateix. És per un estudi de psicologia.

EN: You can type it yourself. It's for a psychology study.

ESCORT

This favor consists in getting somebody to walk you to the gym Marítim. Each time you get somebody to agree, walk with them until you are close enough to the entrance of the gym so that Experimenter 1 can **see you** and **acknowledges your presence with your guide** (by waving or nodding in your direction). If the experimenter is looking somewhere else, just call her name to attract her attention. When you have parted away from your guide, please meet **Experimenter 1**. If you approach somebody and they say they do not know where the gym is, this counts as a ‘No.’

First, you will have to ask potential helpers if they know where that gym is:

SP: Disculpa, ¿sabes dónde está el gimnasio Marítim?

CT: Disculpi, sap on és el gimnàs Marítim?

EN: Excuse me, do you know where the Maritim Gym is?

In case they say “yes”, you will have to ask them to walk you there:

SP: He quedado allí con una amiga pero no sé llegar. ¿Podría acompañarme hasta allí?

CT: He quedat allà amb una amiga però no se com arribar-hi. Em podria acompanyar fins allà?

EN: I have to meet up with a friend there but I cannot find it. Could you walk me there?

In case they start giving you directions, you have to tell them:

SP: Sí, acabo de estar por allí pero no lo he encontrado. ¿Podría acompañarme hasta allí?

CT: Sí, acabo de passar per allà però no l'he trobat. Em podria acompanyar fins allà?

EN: Yes, I have just been there but I have not found it. Could you walk me there?

PICTURE

This favor consists in painting a moustache on someone's face and getting a picture of him or her. You will have to show the photo to the experimenter as a proof that you have got a “yes”.

When you approach somebody, you will have to follow the following script:

SPANISH: Disculpa, ¿podrías hacerme un favor?

CATALAN: Disculpa, podries fer-me un favor?

ENGLISH: Excuse me, could you do me a favor?

If the answer is “yes”, then you have to tell them:

SP: Tengo que pintarte un bigote con esta pintura [se la enseñas] y hacerte una foto. Sólo la usaremos para un estudio de psicología.

CT: T'he de pintar un bigoti amb aquesta pintura [l'hi ensenyes] i fer-te una foto. Només la farem servir per un estudi de psicología.

EN: I need to paint a moustache on your face with this paint [show it to the person] and take a picture of you. We will only use it for a psychology study.

If they press you for details, you have to offer minimal information:

SP: No la publicaremos en ningún sitio y la borraremos al terminar el estudio.

CT: No la publicarem enlloc i la borrarem en terminar l'estudi.

EN: We won't publish it and we will delete it as soon as we finish the study.

To begin the experiment, please, go to Experimenter 1, who is sitting at the café of the CEM Marítim (see the attached map). Once there, meet the experimenter, who will be holding a red folder from UPF and a red backpack. The experimenter will give you some further instructions. Once you have finished, please, meet the experimenter at that spot again in order to complete a short post-experiment questionnaire and to submit the record of your help requests.

Thank you again and have fun!

The Research Team

[On the next page]

Trial Number	Request Type	Compliance (yes / no)
1	Phone	Y / N
2	Picture	Y / N
3	Escort	Y / N
4		Y / N
5		Y / N

Same as before, the results chart was filling the full sheet.

Post-experiment Survey

1. Taking into account the experiences you just had, imagine you have to ask the 'phone' favor again, until you obtain three successes (that is, 3 people will have agreed to text the experimenter on your behalf). How many people do you think you will need to approach? (Cross the number that corresponds to your best guess)

		3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70

71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100+

2. What is your age? _____

3. What is your gender? (Cross the correct response)

Male

Female

4. What is your degree (grau)? _____

5. What is your nationality? _____