

Supplementary information

Link to the repository with the scripts of the experiment's analysis and html: https://github.com/chris-gabaldon/Non_Parametric_ADO/tree/main/Algorithm

Details of our online experiment

The experiment was designed to obtain a monotonically decreasing response curve. During each experimental session, the participant had to complete a total of 12 blocks of 20 trials each, excluding the last one, which contained only 10 trials. Therefore, the experiment ended after a total of 250 trials. In between blocks, the participant was advised to take a 2 minute break. Additionally, throughout the experimental session, in five randomly chosen trials, attention check questions were displayed so as to verify that the participant was paying attention. An example of these questions, where the participant was asked to choose the correct option out of two statements, can be seen in figure SI 1: “4 is greater than 18” or “18 is greater than 4”. The general structure of the experiment can be observed in Figure SI 1.

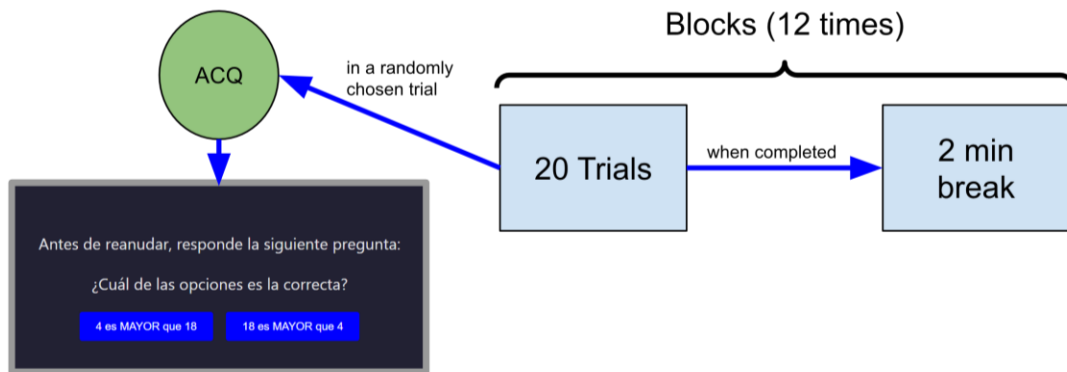


Figure SI 1: General structure of the experiment, and an example of a real attention check question from the experiment asking the participant to select which of the two statements was correct: “4 is greater than 18” or “18 is greater than 4”.

Before beginning with the main task, each participant had to accept the terms and conditions of the study. Then, brief instructions with examples were presented, detailing both the dynamics of each trial and the general structure of the experiment. Since the experiment, as we will show, implied estimating averages, it was also made clear what an average is, and it was emphasised that the participant had to aim to give quick and intuitive answers, without explicit calculations. In addition, it was explicitly stated that the use of calculators or other external aid was prohibited.

The design parameter d to be optimized was then the true average of a list of 8 numbers that would then be displayed. We defined 12 d values: 20, 25, 35, 41, 44, 47, 53, 56, 59, 65, 75, and 80. In each trial, the algorithm (the ADOchoose function) selected one of the proposed designs.

During the trial, the 8 numbers were presented one by one sequentially during 200ms each, with intervals of 300ms between numbers, and at the end, the participant had to decide whether the average was greater or less than 50 by clicking one of the two buttons on the screen indicating these options. This answer was then used by the selection algorithm (using the function $exp.update(d', y)$ described in the section *Open-source Python Package - Installation and Usage*) to determine the design of the next

trial. Figure SI 2 depicts a detailed outline of the development of a typical trial.

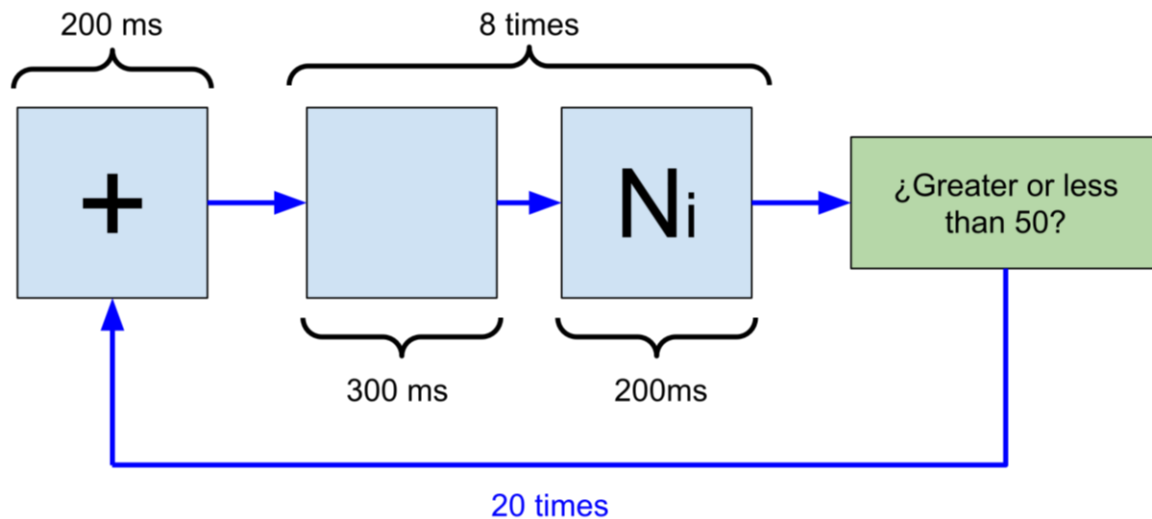


Figura SI 2: General structure of a block of 20 trials. Each trial started with a fixation cross “+” shown during 200ms. Then, eight numbers (based on a true average chosen by the algorithm) were displayed on screen during 200ms (with intervals of 300ms between numbers). Subsequently, the participant had to decide whether the average of the numbers was greater or less than 50.

For example, if design X, corresponding to a target average of 53, was selected, 8 integers were randomly generated whose mean was 53 and whose standard deviation was around 15 (for more technical details, see the corresponding function in the experiment’s script). This step added an additional component of difficulty: a smaller standard deviation would imply that the numbers are similar to each other, which would make it easier to decide whether the average is greater or smaller than 50. Conversely, with greater dispersion, the task becomes more ambiguous and challenging.

A possible example of a sequence of 8 numbers with average 53 and standard deviation ≈ 15 is: [65, 53, 68, 30, 64, 32, 43, 69].

At the end of the experiment, participants had to fill in a form with their age range, educational level and gender, in order to collect information that could be needed for future analysis.

The following images are screenshots of the experiment:

Bienvenido/a al Experimento de percepción numérica

Laboratorio de Neurociencias de la Universidad Torcuato Di Tella

Gracias por elegir participar en nuestro experimento online. Tu participación es voluntaria y podrás abortarlo en cualquier momento. Todos los datos serán anónimos y utilizados exclusivamente con fines de investigación. Al presionar en "Acepto las condiciones" estarás confirmando que:

- Tenés más de 18 años.
- Aceptas participar en el estudio.
- Te comprometes a seguir las instrucciones del experimento.

Considera que la duración aproximada del experimento es de 40 minutos.

☐ Acepto las condiciones.

Instrucciones del experimento de percepción numérica

Por favor lee con atención:

- El experimento evaluará tu capacidad para **estimar promedios** de conjuntos de números. El promedio se calcula sumando todos los números y dividiendo el resultado entre la cantidad de números (por ejemplo, el promedio de 10 y 4 es 7).
- El experimento consiste en 10 bloques, cada uno con 20 tareas, sumando un total de **200 tareas**.
- En cada tarea, se te mostrará una serie de números en el centro de la pantalla (**de a uno a la vez**). Luego, deberás indicar presionando un botón, si el promedio de esos números es **mayor o menor** que 50.
- Luego de responder, el botón se tornará de color **celeste**, indicando que se guardó tu respuesta. Después, deberás presionar un botón que indica "Reanudar", lo cual te permitirá descansar entre cada tarea.
- Se espera que **respondas intuitivamente**, sin usar calculadoras ni realizar cálculos precisos.
- Se recomienda tomar **al menos un minuto de descanso** entre bloques de 20 tareas para evitar la fatiga y mejorar tu rendimiento.
- Al finalizar el experimento podrás escribir tu correo electrónico para que nos pongamos en contacto para entregarte un obsequio 📁.

Presiona el botón "**Comenzar**" para iniciar el primer bloque de tareas. Asegúrate de estar en un lugar tranquilo y recordá que el experimento dura aproximadamente 40 minutos.

¡Mucha suerte!

[Comenzar](#)

Tarea 1/200

+

Tarea 2/200

68

Tarea 2/200

Promedio MENOR a 50

Promedio MAYOR a 50

Gracias por participar en el experimento.

Laboratorio de Neurociencias de la Universidad Torcuato Di Tella

¡Respondiste correctamente 0 tareas!

En el transcurso de los próximos días, nos pondremos en contacto para que puedas retirar un obsequio por haber finalizado el experimento.

Contactate con nosotros por cualquier inquietud

Universidad Torcuato Di Tella. Av. Figueroa Alcorta 7350, Ciudad Autónoma de Buenos Aires, Argentina

email: c_gabaldon@outlook.com

Si querés obtener más información acerca de nosotros, presiona acá para visitar nuestra página web.

Initial version of the experiment:

In a first version of the online experiment, ran to calibrate the experimental details, a different set of designs (averages of the presented lists of numbers) was used, different from the one detailed in the section above. In this case, 10 values were used instead of 12: $d = 41, 44, 46, 48, 49, 51, 52, 54, 56, 59$.

The experimental routine was the same as the one used in the final version, although the number of participants assigned to each design selection algorithm varied. A total of 31 participants took part, distributed as follows: 10 used the ADO algorithm with a normal-informative (NI) prior, 9 with a Gaussian prior, and the remaining 12 performed the experiment with random selection. All these participants performed 200 trials (50 less than in the final version).

Due to the higher concentration of designs close to the fixed value 50, the overall difficulty of the experiment increased. This was reflected in a higher variability in the individual responses, and in

several individual cases the monotonicity condition was lost. The average response curve obtained is shown in Figure SI 3. Compared to the corresponding figure of the final experiment (Figure 17 in the main article), a higher dispersion (larger error bars) and a lower saturation at the extremes can be observed.

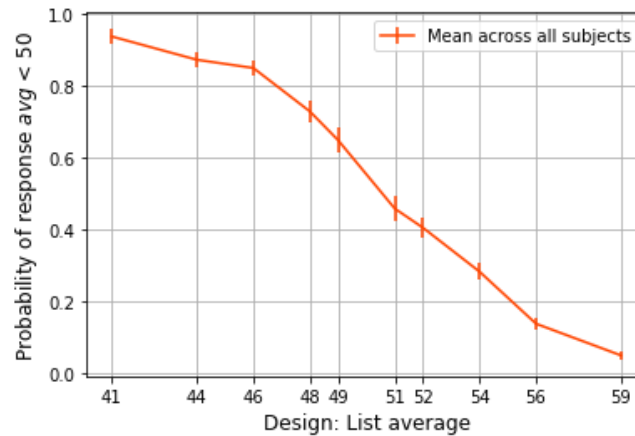


Figure SI 3: average curve and standard errors among all participants, with 200 trials per participant.

A consequence of the low saturation at the extremes can be seen in the distribution of the chosen designs, Figure SI 4. Compared with the same histogram for the final experiment (Figure 19 in the main text), it can be seen that in this first version of the experiment only one design point ($d = 41$ in Figure SI 4 below) is selected fewer times when compared with the other designs.

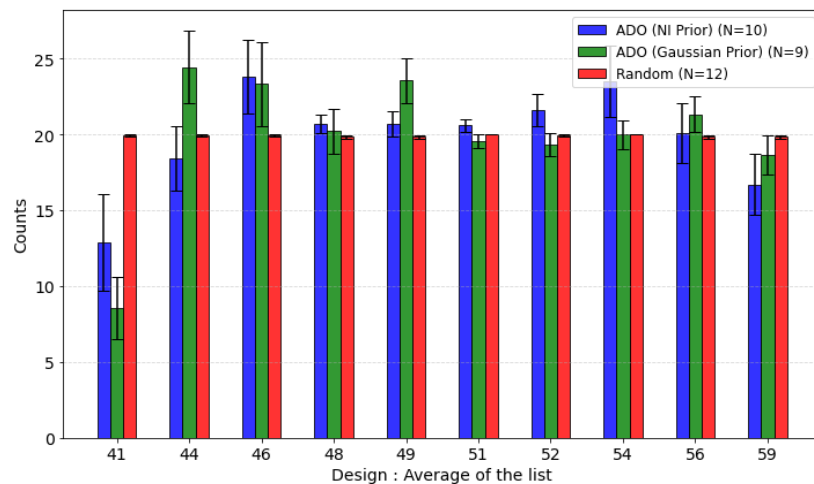


Figura SI 4: Distribution of selected designs for the three experimental groups in the first version of the experiment.

The same MSE analysis as developed in Section "*Experiment*" of the main article was performed, and the curves in this case are shown in Figure SI 5. As expected, since the measured curves do not saturate, there is no clear advantage in the use of the ADO algorithms over the random one. Because of that we redesigned the second version of the experiment with more saturated design points, where we observed what we expected: a higher saturation, accompanied by a faster convergence of the adaptive algorithms compared to the random strategy.

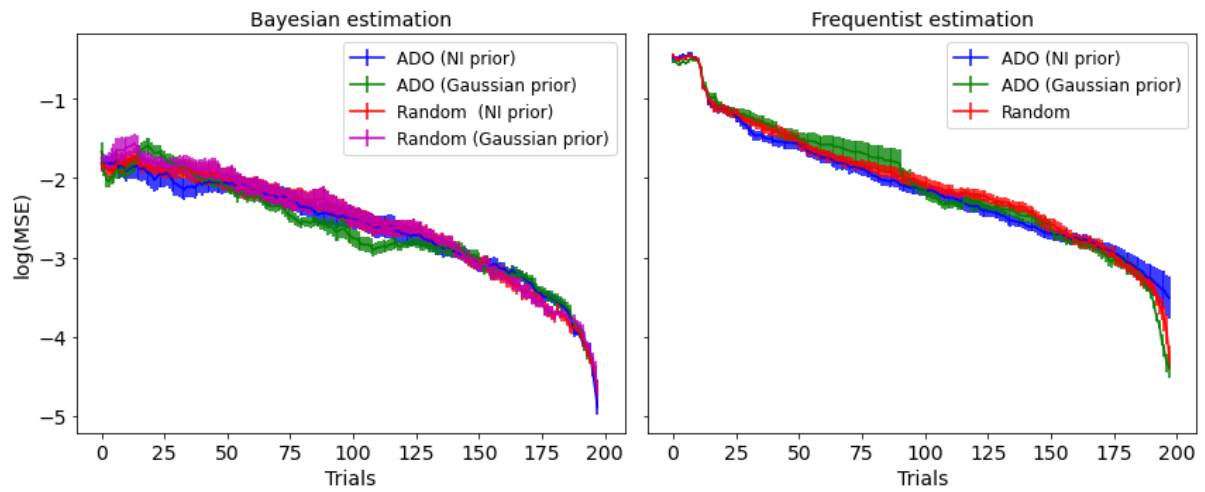


Figura SI 5: participants mean MSEs and its associated standard error as a function of the trial number using ADO or a random strategy for data collection and Bayesian (left) or frequentist (right) inference. Results using our nonparametric ADO algorithm or a random strategy are shown.