# Experimental economics

## Lecture 5: Conducting the experiment

Matej Lorko
matej.lorko@euba.sk
Materials: www.lorko.sk/lectures

References:

- Weimann, J., & Brosig-Koch, J. (2019). Methods in experimental economics. Springer International Publishing. Chicago
- Jacquemet, N., & l'Haridon, O. (2018). Experimental economics. Cambridge University Press.

# Choosing the Design and the Treatments

- The choice of the experimental design depends on the specific research question to be answered by the experiment. Once this question has been formulated, the experiment must be designed in such a way that it produces data that make it possible to decide how the research question can be answered. Ideally, this is achieved by deriving hypotheses from the research question. The hypotheses can then be either confirmed or rejected on the basis of the experimental data. Formulating hypotheses thus serves, above all, the purpose of determining the experimental design.

- All those who are conducting an experiment should also ask themselves what information is being provided by the data generated in the experiment. On the basis of this information, are they really in a position to decide whether or not a hypothesis should be rejected and whether it can be clearly separated from other hypotheses? Only if these questions can be answered with “yes” is the design tailored to the research question.

- When formulating the research question, it is not only the creativity of the researcher that is decisive, but also a thorough investigation of the scientific world. Experimental research is directly related to economic theory. It is therefore also important to ascertain whether there are models in the literature that are relevant to the research question under consideration.

- For economists, this question is of particular significance because there is a reference point of sorts for the interpretation of experimental results that can rarely be avoided: What is the prediction that can be derived from rational (and self-interested) behavior? In order to be able to answer this question, a suitable model is required. Either this already exists in the theoretical literature or it needs to be developed and solved.

- The situation is somewhat different when we are dealing with an experiment in which it is already clear that rational and self-interested behavior does not provide any useful predictions. This leaves us with two possibilities. Either we have a model that deviates from these assumptions and tries to organize the experimental findings, or we limit ourselves to a purely exploratory study that tries to gain information about individual behavior that might help to find an explanation for what happens in the experiment, for example, by using hypotheses from psychology.

# Choosing the Design and the Treatments

- When it comes to the specific design of the experiment, the key question is what needs to be controlled and how this should be achieved in each individual case. Basically there are four things that can (and should) be controlled.

- Preferences, Motives, Attitudes: When analyzing behavioral motives that deviate from pure payoff maximization, it is important to realize that such motives cannot be observed directly. This means that their existence can only be concluded if they lead to deviations from payoff-maximizing behavior. This is an important point to consider when designing the experiment. If it is to be possible to deduce certain behavioral motives from the subjects' behavior, then the monetary incentives must be set in such a way that a specific motive can be deduced as clearly as possible from the deviations from purely selfish behavior.

- Constraints under which decisions are to be made: there are two important areas that can be designed in practically every experiment: first, the payoff conditions and, second, the information that the subjects receive.

- The manner of presentation (the frame): every experimenter must realize that it really is necessary to make an active choice, because there is no such thing as an experiment without a frame. The second issue to be settled is whether one prefers as neutral a presentation of the decision problem as possible or whether the aim is to approximate a frame as it is actually found in the real world

- Experience and prior knowledge of the subjects of the experiment: People's prior knowledge or experience can systematically influence their behavior. If these factors are not controlled, there is a risk of having selection effects in the experiment and these should be avoided if possible. If in one treatment mainly economists participate and in another, mostly humanities students, this can lead to a difference that looks like an treatment effect, but which in reality can have other causes. Also, wehther or not subjects had an experience with a similar experiment (or laboratory experiment in general) can make a difference.

- Finally a decision must be made as to whether a within-subject or between-subject design is to be used. Also, it is necessary to determine whether the data provided by the experiment will conform to the statistical requirements that must be fulfilled for a meaningful analysis.

# Creating the Instructions

- The subjects need to be informed about the course of the experiment and this is done with the help of instructions given to them.
- Of course, there is no authoritative standard text, but in our experience it has proved useful to introduce the instructions by briefly informing the subjects in the experiment that they can earn money through their participation and whether it depends both on their own actions and those of other subjects in the experiment how much money they are paid in the end. It should also be emphasized that leaving the workstation and talking to other subjects during the experiment is prohibited.
- If the experiment involves communication between the subjects, this must of course be explained separately. How to get the experimenter's attention to ask questions, how long the experiment takes, whether there is a show-up fee and – if the experiment consists of several parts – how many parts the experiment has and how these parts are related are also typically explained in the instructions.
- After this general information has been provided, it is time to describe the experimental design. It is important that this is done in such a way that every subject understands exactly what decision he has to make and what consequences this decision has for him and eventually for the other players.
- However, a caveat needs to be made in this connection. Particularly in experiments in which learning behavior is to be investigated, it is sometimes necessary not to tell the subjects everything that will happen. If they knew everything, there would be nothing left to learn. It must nonetheless be ensured that the subjects do not receive untrue information.
- Instructions should be as simple as possible and not too long. The longer the text, the more likely the subjects will not read it to the end.

# Instructions and Comprehension Tests

- All the elements of an experimental design must be communicated to the subjects of the experiment. This is done in the instructions, which are either provided verbally or distributed in writing to the subjects. Two important questions are of interest here. First, how can the instructions be conveyed in such a way that it is certain that all the subjects have actually taken note of and understood them, and second, how can potentially distracting effects be eliminated?

- Ideally, instructions should be in writing and distributed as a document to the subjects. An important reason for this is that it is then certain that the subjects can look at the instructions again during the ongoing experiment if anything is unclear to them. This also rules out variations in the presentation of the instructions from session to session that can undoubtedly take place if the instructions are communicated verbally (even if arises simply through a variation in the emphasis of some words).

- However, by providing the instructions verbally, it is possible to ensure that they are common knowledge for all the subjects. In other words, the subjects know that everyone knows that everyone knows... that everyone knows what is in the instructions. It is therefore not at all unusual for the instructions to be distributed in writing, and also to be read out.

- As far as the content of the instructions is concerned, there are three points to bear in mind: (1) The description of the experiment should be as short and concise as possible. (2) The description of the experiment must be as simple and understandable as possible. (3) Instructions are the point where experimenter demand effects could be generated or norms might be triggered. This is something to be aware of, i.e. when writing the instructions it is important to remember that signals are being sent to the subjects who could possibly use them to interpret what they should do.

# Instructions and Comprehension Tests

- How should we deal with questions that the subjects still have after they have received the instructions? We recommend that questions not be asked publicly. For this reason, reading the instructions out loud should not be concluded by asking the group if anyone has a question but rather by pointing out that questions can only be asked in the strictest confidence and then answered one-on-one between the subject and the experimenter.

- Why is it not advisable to have questions asked publicly? The problem is that there is no control over what is asked. As a consequence, questions might be asked that are not about understanding the experiment, but rather about giving an indication of indi- vidual expectations or behavior or how one should behave.

- The saying that trust is good but control is better also applies to experimenters. It is therefore a good idea to check whether the subjects really have understood the experiment. Control questions are therefore important, but they also entail the risks already mentioned. They can trigger experimenter demand effects, activate norms or lead to anchoring effects.

- In any case, all the subjects should be given the *same* control questions. This means that if values are determined randomly, then this should be done once for all the subjects and not for each one individually. This ensures that the group of subjects is homogeneous in terms of subjects' previous experience.

# The Experimenter Demand Effect

- The experimenter influences what happens in an experiment through different channels. Some are obvious, such as the instructions given to the subjects by the experimenter, or the exercises used to test whether the subjects have understood the experiment. Others are less obvious, but just as important. Thus, the experimenter can consciously or unconsciously exert social pressure or certain expectations can be generated in the subjects as to the purpose of the experiment and what behavior is now expected of them.

- In laboratory experiments the interaction between the experimenter and the subject is inevitable (even if it is through the design developed by the experimenter). It cannot therefore be a question of avoiding any kind of interaction, but rather of designing it in such a way that it does not lead to any distorting influence on the behavior of the subjects (experimenter demand effect), thereby curtailing the interpretability of the data obtained.

- Cognitive experimenter demand effects occur because the experimenter has to explain the experiment to the subjects. Understanding this explanation is a cognitive process and it may well happen that how it is explained leads to it being understood in a particular way, for example what is appropriate behavior in the experimental situation. Experimenters should be aware of the fact that subjects may take every word seriously and, therefore, that every word used by the experimenter should be carefully considered.

- In addition to cognitive experimenter demand effects, undesirable manipulation of the subjects may also result from social pressure, which can arise both between the subjects and vertically from the experimenter. There are many reasons why people succumb to social pressure. For example, a role may be played by the desire for conformity, or by social acceptance, which is experienced when acting in accordance with a social norm. It is quite possible that there are also subjects who attach great importance to being nonconformist and therefore oppose any social pressure. While it may not be too bold a hypothesis to suggest that nonconformists are rare, the widespread desire to conform to social norms is well known.

# The Experimenter Demand Effect

- The instructions that the subjects receive at the beginning of an experiment are ideally suited to creating massive experimenter demand effects. The language used, for example, is suspected of doing this. It is possible to describe things in an emphatically neutral way or to “load” them with valuations to a greater or lesser extent.

- Liberman et al. (2004) report on two public good experiments, which were identical except for the names of the games provided to the subjects. One was a “Community Game” and the other was a “Wall Street Game”. The names actually had a huge influence on the results, with much more cooperation in the Community Game than in the Wall Street Game.

- In the experiment by Burnham et al. (2000), too, altering only one word triggered substantial effects. In their experiment, two players could significantly increase their payoffs compared to the equilibrium payoff if player 1 trusted player 2 and player 2 acted reciprocally, thus vindicating the trust. In the first treatment, the other player was called the “partner”, while in the second treatment the word “opponent” was used. The word “partner” led to significantly more trust and trustworthiness at the beginning of the experiment. Admittedly, both declined in later rounds.

- The decisive question in both cases is what effect is actually present. Is it a particular value judgment associated with the respective terms, or is it an experimenter demand effect? In the latter case, when a game is called “Wall Street Game”, the subjects might have the feeling that the experimenter wants to test how well they can assert themselves. If the game is called “Community Game”, the experimenter might want to know how well the subjects perform as social beings. If the other player is called a partner, the experimenter apparently wants to test the ability to cooperate. If, on the other hand, the other player is designated an “opponent”, then competition is evidently at issue and it is a matter of asserting oneself.

- Experimenter demand effects can act in different directions. The reference point is the experimental effect expected in the experiment. The experimenter demand effect may be in the same direction, opposite or orthogonal to the experimental effect.

- The most problematic is the experimenter demand effect that acts in the same direction as the expected experimental effect. In such a case, it is difficult to decide whether what is observed is due to the experimenter demand effect or to the experimental conditions. If the experimenter demand effect runs in the opposite direction, it can just offset the experimental effect and no clear effects can be detected. The least problematic are experimenter demand effects that are orthogonal to the experiment effect. They may not influence the behavior of the subjects in a way that hinders the interpretation of the results of the experiment.

# The Frame of the Experiment

- The frame of an experiment is the way in which a specific decision problem is presented to the subjects. Framing effects are the changes in the subjects’ behavior that occur solely because the presentation of the decision problem is varied without changing the problem itself and its solution.

- In the recent literature, two types of framing effects play a special role. The first occurs when only the name of a game is changed (label frame). We have already referred to the following example in the previous section. Whether you call a public good experiment “Community Game” or “Wall Street Game” makes a major difference.

- The second framing effect that has attracted much attention is what is named the valence frame. This means that certain terms are loaded with respect to the values or preconceptions associated with them. The standard example again concerns the public good game, which can be played in a “Give” or a “Take” treatment (Dufwenberg et al. 2011).

- In the Give frame, the individual members of a group each receive an initial endowment (zi), which they can either keep or pass on to any part of a joint project (the public good). In the Take frame, the entire initial endowment (i.e. the sum of the zi) is in the joint project and the subjects can withdraw money up to the amount of zi. Obviously the same decision problem is involved in both cases, but the experimental findings show that significantly more is invested in the public project under the Give frame than under the Take frame.

- The observation that the results of experiments can be strongly influenced by the respective frame has led to the emergence of neutral frames as a standard – at least when it comes to testing general models. This means that names that could be given to an interaction or the persons involved are consciously avoided and that the description of the experiment is designed as value-free and neutral as possible.

# The Frame of the Experiment

- Let us assume that when subjects enter a laboratory and receive instructions for an experiment, they first try to understand what the experiment is about and what behavior is expected from them. The frames of the experiment then serve as an orientation aid for the subjects. What is the name of the experiment? What is the name of the activity I need to perform? What conclusions can be drawn from the type of task I am faced with here?

- Questions of this kind will occupy the subjects. It should be borne in mind that the subjects assume that the frame – i.e. the answers to their questions – was set by the experimenter. The person who wrote the instructions and designed the experiment thus provides the information that the subjects use to make sense of the experiment.

- This means that each frame – no matter how it is designed – is always associated with a potential experimenter demand effect. If one accepts this consideration, the question of whether a change of the frame impacts on the behavior can also depend on whether this alters the potential experimenter demand effect and whether this in turn has any impact.

- Of course, the behavior of subjects is not only determined by experimenter demand effects. Ideally, their influence is rather small and the effect of monetary incentives dominates the decision. Nevertheless, when designing an experiment, one should at least be aware of the potential connection between frames and experimenter demand effects.

- The second way the frame has an impact is that it can also influence the beliefs of the subjects about other subjects' behavior. This is all the more the case because the frame directly creates common knowledge.

- A third way it impacts arises because a frame can be accompanied by the activation of social norms. It is important to note that such norms can also have an influence in the real world. If a real phenomenon is to be simulated in the laboratory, a corresponding frame should therefore be included.

# Double-Blind Design

- Double-blind procedure is an experimental design that ensures that the experimenters cannot observe how the individual subject acts and that also maintains anonymity between the subjects. This is generally achieved by having the subjects drawing identification numbers randomly and in a concealed manner. As a result, the experimenters know how, for example, subject number 17 behaved, but not who number 17 is. A single-blind design means that the subjects cannot observe each other, but the experimenter sees what the individual person is doing.

- It is essential to see double-blind designs in close conjunction with the experimenter demand effect. This is necessary because it cannot be ruled out that the use of a double-blind design itself will trigger an experimenter demand effect. If experimenters explicitly draw the attention of their subjects to the fact that they are acting anonymously and cannot be observed by the experimenter, then it is obvious that the subjects will think about why it is so important to the experimenter that they can act without being observed. Therefore, when using a double-blind design, it is not advisable to explicitly point out that this is intended to achieve anonymity.

- Double-blind designs are particularly effective where a strong experimenter demand effect is expected, however, if a sufficiently high degree of anonymity is already guaranteed by a single-blid design, double-blind might not be necessary.

# Setting Up an Experimental Laboratory

- The basic arrangement of a laboratory consists of a series of computer workstations for the subjects and a workstation for the experimenter who manages and conducts the experiment. There are two aspects which must be taken into account here and which are to a certain extent contradictory.

- On the one hand, it is necessary for the experimenter to be able to monitor the subjects of the experiment, for example, to prevent unwanted communication. On the other hand, it is important to avoid as far as possible the subjects feeling that they are under observation.

- Every laboratory needs at least two types of software. One that can be used to program the experiments so that they can be run over a computer network, and one that can be used for the purpose of administration and recruitment of the subjects.

- z-Tree has become the global standard for experiment programming. The tool was developed by Urs Fischbacher and has been updated and further developed for many years. z-Tree offers the possibility to program almost any experiment in a relatively simple way. Since it is precisely adapted to the needs of experimental economic research, it mainly contains elements that are frequently used there. This has the advantage that z-Tree is relatively streamlined and therefore easy to learn. The program is available for free and it is well documented.

- In addition to the programming of experiments, laboratories need professional recruitment and supervision of subjects if these laboratories are to conduct experiments on a regular basis. Programs are also available for this purpose. ORSEE, developed by Ben Greiner,5 played a similar role to z-Tree for quite some time. In contrast to the programming of the experiments, when it comes to the recruitment software it is not so important that many laboratories choose the same program, since the recruitment always takes place locally.

- This makes it a little easier for newcomers and is probably the reason why HROOT has in the meantime become a strong competitor for ORSEE. Both solutions have a similar scope of services. With the aid of the recruitment software, it is possible to manage all the potential subjects online. People who would like to participate in experiments can register for the database online. The most important characteristics of the respective person are recorded in the database. In addition to demographic data, this includes above all information on the experiments in which a particular person has already participated. It is very important to know this because, as a rule, researchers are interested in people who have not yet had any experience with the planned experimental setup. Sometimes, however, it is desirable to invite precisely those people who have already participated in a similar experiment to the laboratory.

# Writing the Plan of Procedure

- Once the instructions are written, it may prove useful to create a plan of procedure for the experiment. This is particularly the case when the different sessions and treatments are not always carried out by the same people.

- A plan of procedure is essential to ensure that all the experiments proceed in exactly the same way. This plan should describe as precisely as possible what is to happen during the experiment.

- This begins with the subjects entering the laboratory. Should they be admitted individually or as a group? What measures must be taken to maintain anonymity? How are the instructions distributed or read aloud? What is the procedure for responding to questions from the subjects?

- It is vital that the plan of procedure describes all these details so that each and every person who conducts the experiment knows exactly what to do, how to do and when to do it, from the admission of the subjects to the final payment of the payoffs of the experiment.

- Creating a plan of procedure has another advantage: it facilitates the replication of the experiment.

# The Pilot Experiment

- Once the plan of procedure has been drawn up and all the detailed issues described in it have been addressed, the experiment could in principle commence. But before doing so, it is often wise to run a pilot experiment. The purpose of such a pilot is to check whether everything runs exactly as imagined. An important point here, of course, is the software or the specific program that was written to conduct the experiment. Does it perform under realistic conditions – even if the users make mistakes while entering their data (as subjects sometimes do)? It is much more unpleasant to discover an error during the actual experiment than during a pilot experiment.

- If the pilot experiment is to be used purely for testing the processes and the software, it can be run with people who know that it is a pilot experiment. If, however, the aim is to gather valid data in the pilot experiment, there should be no deviation from the actual experiment when selecting the subjects, i.e. the same recruitment method and the same number of subjects must be used. Furthermore, the payoffs need to be real and equal to the payoffs of the planned experiment.

- In addition to the software, the instructions should also be thoroughly checked in a pilot experiment. After the experiment, the subjects can be informed that they were involved in a pilot experiment and asked how easy it was for them to understand the instructions and how well they understood them.

- After the completion of a pilot experiment and the evaluation of its results, the question arises as to how to deal with the data that was obtained. If the subjects were selected and paid off as they would be in the experiment, if everything ran smoothly and if no changes to the design or the way the experiment was carried out were necessary, there is nothing against integrating the data into the data set of the experiment. The pilot experiment therefore does not differ in any way from the other sessions in which the experiment is conducted.

# Manipulation check

- In both economics and psychology, a manipulation check serves as a crucial step in ensuring that an experimental treatment worked as intended. It assesses whether participants actually perceived, understood, or experienced the manipulation in the way the researcher designed. Without this verification, it is difficult to determine whether observed differences in behavior or attitudes truly result from the experimental condition or from unrelated factors such as confusion, inattentiveness, or noise.

- A successful manipulation check provides evidence for internal validity, confirming that the independent variable exerted the intended influence. When a manipulation fails—meaning that participants do not detect or respond to the intended change—the interpretation of treatment effects becomes ambiguous. For this reason, manipulation checks are often included immediately after the main task but before debriefing, to avoid influencing participant behavior during the experiment itself.

- Manipulation checks can take several forms. Direct checks typically involve asking participants about their perceptions or beliefs related to the manipulation (for instance, "How fair did you find the allocation process?"). Indirect checks, in contrast, rely on behavioral or physiological indicators that indirectly reflect the manipulation's impact—such as reaction times, emotional tone, or contribution levels.

- The role and form of manipulation checks vary across disciplines. In psychology, manipulations often target cognitive or emotional states, such as mood, trust, or perceived fairness, so checks frequently measure subjective perception. In experimental economics, manipulations more often concern incentives, information structures, or institutional rules; thus, checks typically test whether participants understood the rules, incentives, or timing of decisions. In both fields, the manipulation check strengthens credibility by demonstrating that treatments were both salient and correctly interpreted.

# Recruiting the Subjects

- Before an experiment can be carried out, it is essential to confirm that suitable subjects are available. Recruitment is relatively easy if it is limited to students as subjects. Ideally, the university administration is cooperative and allows the laboratory, for example, to write to first-year students by email informing them of the laboratory, the possibilities of earning money and the registration procedure. If there is no possibility to send electronic mail to the potential subjects, it is necessary to take the more difficult path and go through the lecture theaters to introduce the laboratory.

- If the recruitment was successful, the laboratory possesses a pool of potential subjects for selecting those to be invited after the pilot experiment. The criteria used to do this can be very diverse, but it is crucial that they always take into account a principle that must be observed when inviting subjects: selection bias is to be avoided.

- For this purpose, it is necessary, for example, for the subjects to be randomly assigned to the different experimental treatments. The software used for the invitations is designed to do this, using a random selection procedure to choose the people to be invited for each treatment.

- It is advisable always to invite a few people as substitutes, who only participate in the experiment if registered subjects do not show up. When inviting the subjects, it is important to inform them that they may act as a substitute and will therefore only be used if necessary. It is also important that the substitutes are paid for showing up, even if they are not used.

# Conducting an Experiment

- Once the pilot experiment has been evaluated, all the necessary design adjustments have been made and enough subjects have registered for the experiment, the actual experiment can proceed. The first step, of course, is to get the subjects into the laboratory.

- The question of how the subjects actually get into the laboratory depends largely on the specific experiment. The issue to decide here is how to manage the required level of anonymity between the subjects. If it is essential that the subjects have no opportunity to identify themselves, then it makes little sense to invite them all to the laboratory together. In such cases, a somewhat more complex procedure is required.

- If the anonymity of the subjects is not an important aspect of the experiment, the complicated process of fetching the subjects can be dispensed with and they can simply be sent to a location near the laboratory. This can be a separate room or a corridor. Once everyone is gathered, the substitutes find out whether they can participate or go home after receiving their compensation.

- Two tasks then follow. First, the names of the subjects are checked so that after the experiment the names of those who took part in the experiment and of those who may have been absent without an excuse can be entered in the subject database. The second task is to assign the subjects to the various roles. In most experiments, there are different roles: buyers or sellers, proposers or receivers and so on.

- Although it is not uncommon for there to be only one role, for example in the provision of public goods, the experiment is still run in several groups, so the groups have to be made up. It makes good sense to combine the two tasks. When the names are checked, the subjects draw “lots” that randomly assign them to a role or group. A well-organized laboratory holds suitable objects, such as table tennis balls, wooden balls or the like, that can be used as lots.

- Drawing lots for roles and group memberships ensures that the assignment is randomized, which is extremely important in order to avoid selection effects. At the same time, identification numbers can be drawn with the lots. Obviously, this has to be done in such a way that the experimenter cannot see the identification number. When making decisions in the experiment, the subjects can then enter their number instead of their name. This increases the anonymity of the decisions.

# Conducting an Experiment

- There is no set rule as to how the instructions are to be communicated to the subjects. However, it is recommended to first hand them out in writing, printed on a sheet of paper (not online), and then, if possible, read them out loud.

- Reading the text aloud almost always has the effect that the subjects simultaneously read the text on their sheets, thus ensuring that they have read it to the end. If the instructions are not read aloud, this effect is lost and the experimenter can only assume that everyone actually has read everything to the end.

- If all subjects are in the same laboratory room and no special arrangements to ensure anonymity have to be made and if all subjects participate in the same treatment, there is no reason why they should not be called together as a group and the instructions read out. However, reading out instructions should be as homogeneous as possible across sessions and treatments (i.e. ideally the same experimenter should be involved).

- Once all subjects have read the instructions, they should have the opportunity to ask questions. It is better not to have these questions asked publicly, but privately, i.e. in a conversation between the subject and the experimenter.

# Conducting an Experiment

- Once all the decisions have been made and the experiment is over, it is time for the payments to be made to the subjects. Before this can happen, there is occasionally a problem that we would like to discuss briefly. The behavior of the subjects can vary greatly, and this may also manifest itself in the fact that the individual participants in the experiment solve the decision problems at very different speeds. This in turn may mean that individual subjects finish the experiment much earlier than others. What is the best way to handle this?

- If the payment does not depend on the speed at which decisions are made, but only on the decisions themselves, then the earlier one leaves the laboratory, the higher the hourly rate of pay. This creates strong incentives to make decisions as quickly as possible. However, this is not in the interest of the experimenter, because speed can easily be at the expense of care. Subjects should think carefully and very precisely about their decisions and not hastily. Therefore, it should not pay off to be faster than the other subjects in the experiment.

- There is another compelling reason for not making payment until everyone is finished. If somebody were to be paid off while the experiment is in progress, it would inevitably lead to those who are not yet finished being disturbed and having the feeling that they have to hurry, because others can already leave. This should be avoided at all costs. The subjects do not need to know how quickly the other subjects perform their tasks and restlessness in the laboratory is inherently not good for an experiment.

- Once all the subjects have completed the experiment, payment can be made. Ideally, payment should not be made in the same room as the experiment. If this cannot be avoided, it should at least be ensured that the anonymity of the payment is otherwise secured.

# Achieving the best possible control of the subjects' preferences

- **The payoffs should be noticeable.**This means that the subjects can only be expected to pay attention to the payoffs if they are structured in such a way that it is worth paying attention to them.

- **Subjective costs should be minimized**. This means that it should be made as easy as possible for the subjects to understand the task presented in the experiment and to make the best decision for them. This is one of the reasons why experiments should be simple.

- **Use neutral language.** The main aim is to avoid experimenter demand effects. This means that the subjects should not be given the impression that the experiment serves a specific purpose

- **Provide an opportunity to learn.** Even simple games should, however, be practiced by the subjects before the actual experiment takes place. It is quite possible that learning processes take place in the first few rounds of an experiment. If the experiment is not designed to observe these learning processes, then learning the game has no place in the actual experiment. The aim is to test whether the subjects who know and understand the game behave as predicted by the experimental hypothesis or not. Therefore, the learning process must take place before the experiment.

- A very controversial question is whether to conduct an experimental session only once or whether to provide the subjects with a repeated opportunity to experience the experimental situation – for instance, by carrying out the experiment again at intervals of 1 week.

- Repetition tends to increase external validity, since most of the decision problems that are explored in experiments are in fact not one-off occurrences in the real world, but recur at irregular intervals. The disadvantage of repeating the session is that the experimenter has no control over what happens between the repetitions. It may well be that the subjects in the meantime gain experience which has a strong influence on their behavior. The problem is that there is no way of knowing what experience this is. It is presumably this methodological problem that has hitherto prevented experimenters from repeating sessions to any large extent.

# Case study

## Russo-Ukrainian war disinformation:
## the effect of debunking vs. prebunking

Link: https://osf.io/preprints/psyarxiv/w3mfy

Institute of Experimental Psychology - Matej Lorko, Vladimíra Čavojová, Jakub Šrol
Insittute for Financial Policy - Richard Priesol, Paulína Jalakšová, Berenika Tužilová

# Motivation

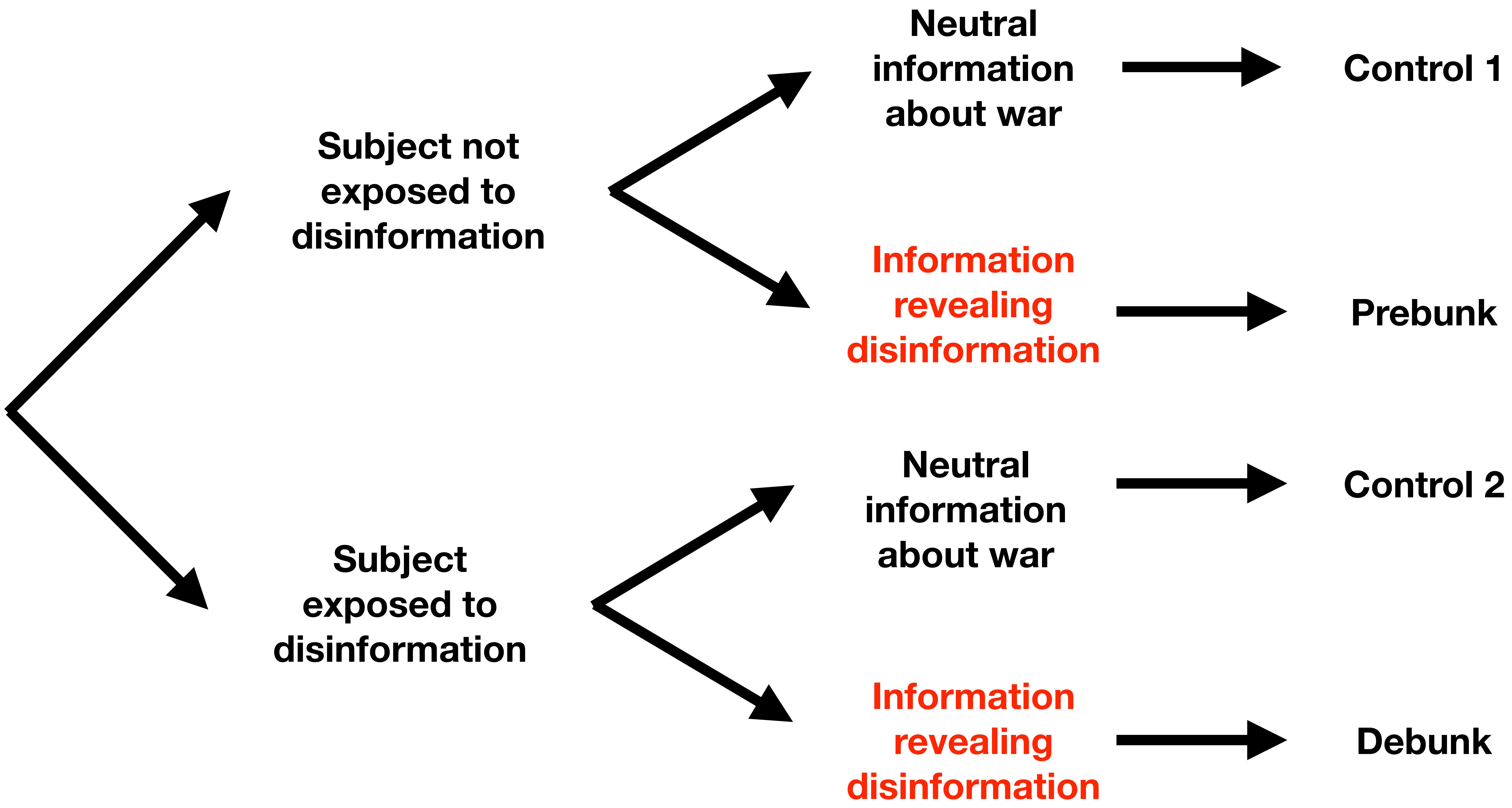
- Russian invasion to Ukraine came with a wave of disinfotmation intended to lower the trust in the motives and actions of Ukraine and Western countries.

- It is therefore essential to search for interventions that could reduce trust in such disinformation.

- The goal of this research is to compare the effectiveness of Debunking vs Prebunking.

- We manipulate the timing at which the corrective message (disinformation refutation) is presented - either **after** the subject is already exposed to disinformation (**debunking**) or **before** it (**prebunking**)

# Treatments


Subject not exposed to disinformation
Subject exposed to disinformation
Neutral information about war
Information revealing disinformation
Neutral information about war
Information revealing disinformation
Control 1
Prebunk
Control 2
Debunk

# Hypotheses

- Hypothesis 1a: Prebunk intervention reduces the trust in disinformation.
- Hypothesis 1b: Debunk intervention reduces the trust in disinformation.
- Hypothesis 2: Debunk intervention reduces the trust in disinformation more than Prebunk intervention.

# Experimental design

- 2 measurements
- Laborarory experiment on student sample (N=220) with 2 sessions, 2 weeks apart
- Survey on representative sample (N=925) with meaurements approximately 30-40 minutes apart

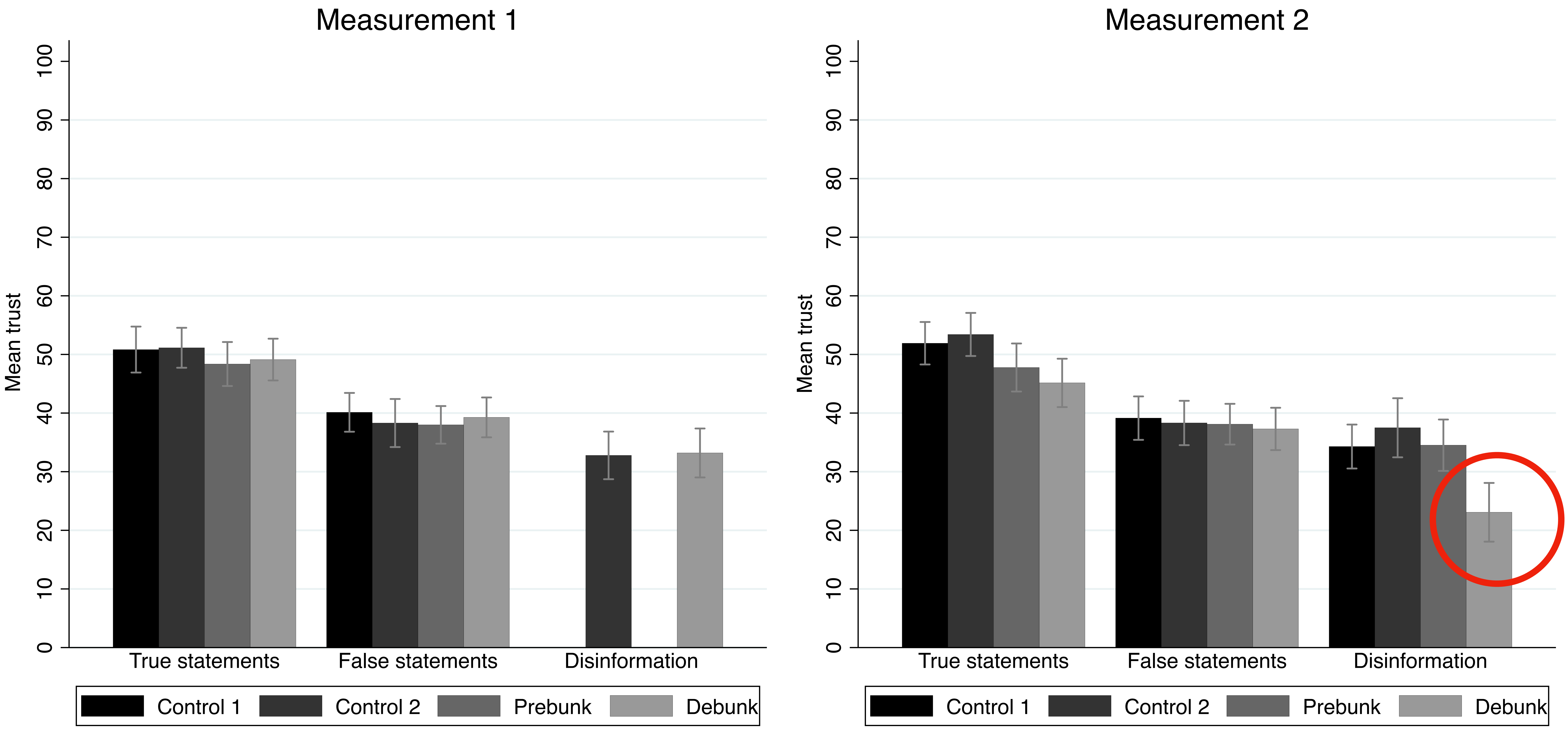
| Treatment | Measurement 1 | Intervention (within Measurement 1) | Measurement 2 | Other measures |
|---|---|---|---|---|
| Control 1 | 10 true information<br>5 false information | Neutral information about war | 10 true information<br>5 false information<br>5 disinformation | · Indication of whether the participant was exposed to those 5 disinformation outside of experiment<br>· Demographics, CRT, etc.<br>· Attituted towards war, Russia, Ukraine, West |
| Control 2 | 10 true information<br>5 false information<br>5 disinformation | Neutral information about war | 10 true information<br>5 false information<br>5 disinformation | |
| Prebunk | 10 true information<br>5 false information | Corrective message | 10 true information<br>5 false information<br>5 disinformation | |
| Debunk | 10 true information<br>5 false information<br>5 disinformation | Corrective message | 10 true information<br>5 false information<br>5 disinformation | |

# Descriptive statistics (Lab)

| | Female | Age | CRT | UKR fan | RUS fan |
|---|---|---|---|---|---|
| Control 1 (N=55) | 71% | 22 (1) | 1.5 | 82% | 5% |
| Control 2 (N=45) | 47% | 23 (4) | 1.7 | 73% | 4% |
| Prebunk (N=60) | 57% | 22 (1) | 1.6 | 87% | 0% |
| Debunk (N=60) | 50% | 22 (2) | 2.0 | 85% | 3% |

| | Blames west | Blames RUS | Blames UKR | Follows news | Avoids news |
|---|---|---|---|---|---|
| Control 1 (N=55) | 22% | 89% | 24% | 11% | 49% |
| Control 2 (N=45) | 16% | 87% | 13% | 16% | 36% |
| Prebunk (N=60) | 25% | 95% | 22% | 17% | 43% |
| Debunk (N=60) | 15% | 92% | 10% | 35% | 47% |

# Results (Lab)


Measurement 1
Measurement 2
Mean trust
0
10
20
30
40
50
60
70
80
90
100
True statements
False statements
Disinformation
Control 1
Control 2
Prebunk
Debunk

# Lab results - trust in war-related information (mean, SD)

| | True 1 | True 2 | T1=T2? | False 1 | False 2 | F1=F2? | Disinfo 1 | Disinfo 2 | D1=D2? |
|---|---|---|---|---|---|---|---|---|---|
| Control 1 (N=55) | 50 (15) | 52 (14) | YES | 41 (13) | 39 (14) | YES | | 34 (14) | |
| Control 2 (N=45) | 51 (12) | 53 (13) | YES | 40 (16) | 38 (13) | YES | 32 (16) | 37 (17) | **NO** |
| Prebunk (N=60) | 48 (16) | 48 (16) | YES | 38 (13) | 38 (14) | YES | | 35 (17) | |
| Debunk (N=60) | 49 (14) | 45 (16) | YES | 39 (13) | 37 (14) | YES | 33 (17) | 23 (20) | **NO** |
| Statistics (ANOVA, Mann-Whitney, t-test) | Not significant | **Debunk weakly significant** | | Not significant | Not significant | | Not significant | **Debunk significant** | |
| Cross comparison ` | True > False > Disinfo | | | | | | | | |
| | True 2 > False 2 > Disinfo 2 | | | | | | | | |

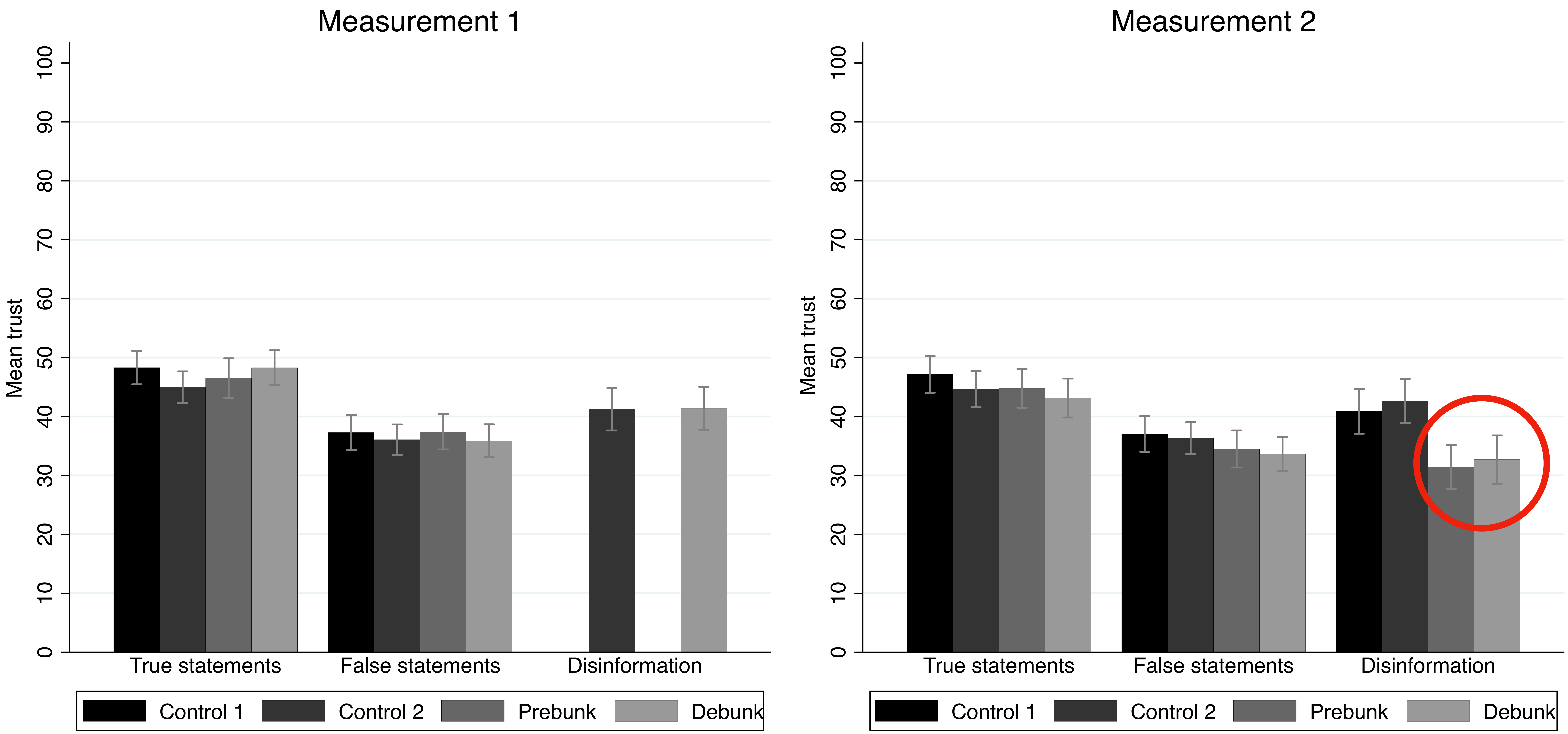
# Laboratory experiment

**Table 3: OLS regressions, Experiment 1 (abbreviated)**

| | (1) True | (2) False | (3) Disinfo | (4) True | (5) False | (6) Disinfo |
|---|---|---|---|---|---|---|
| Corresponding M1 trust | **0.67*** (0.05)** β = .64 | **0.48*** (0.06)** β = .49 | | **0.65*** (0.05)** β = .62 | **0.50*** (0.06)** β = .50 | |
| Prebunk treatment | -3.21 (1.85) β = -.09 | 0.31 (1.97) β = .01 | 0.23 (3.23) β = .01 | -3.44 (1.83) β = -.10 | -0.41 (1.97) β = -.01 | 1.62 (3.29) β = .04 |
| Debunk treatment | **-6.11** (1.85)** β = -.18 | -0.72 (1.97) β = -.02 | **-14.42*** (3.41)** β = -.36 | **-6.98*** (1.83)** β = -.21 | -0.64 (1.97) β = -.02 | **-12.85*** (3.44)** β = -.32 |
| Disinformation included in M1 | | | 3.20 (3.48) β = .09 | | | 4.75 (3.48) β = .13 |
| Controls included | No | No | No | Yes | Yes | Yes |
| Constant | 18.34*** (2.92) | 19.35*** (2.67) | 34.29*** (2.33) | 36.14*** (9.36) | 36.97*** (9.81) | 71.16*** (13.42) |
| N | 220 | 220 | 220 | 220 | 220 | 220 |
| Adjusted $R^2$ | 0.45 | 0.23 | 0.08 | 0.48 | 0.25 | 0.14 |
| | Wald test of the difference between Prebunk and Debunk | | | | | |
| $\chi^2(1)$ | 1.97 | 0.22 | 9.73 | 2.99 | 0.01 | 9.87 |
| p-value | 0.16 | 0.64 | <.01 | 0.09 | 0.92 | <.01 |

*Note.* The table reports OLS regressions predicting trust in true, false, and disinformation statements in M2, without (Columns 1–3) and with (Columns 4–6) control variables. Categorical variables are dummy-coded. For each predictor, unstandardized coefficients, standard errors (in parentheses), and standardized coefficients (β) are shown. Wald tests comparing Prebunk and Debunk treatments appear at the bottom. Statistically significant predictors ($p < .05$) are in bold. Full results are reported in the Supplementary Materials. $p < .05$; p < .01; $p < .001$.

# Results (representative sample)



# Online experiment

**Table 4: OLS regressions, Experiment 2 (abbreviated)**

| | (1) True | (2) False | (3) Disinfo | (4) True | (5) False | (6) Disinfo |
|---|---|---|---|---|---|---|
| Corresponding M1 trust | **0.87*****<br>**(0.02)**<br>β = .81 | **0.78*****<br>**(0.02)**<br>β = .75 | | **0.87*****<br>**(0.02)**<br>β = .81 | **0.78*****<br>**(0.02)**<br>β = .76 | |
| Prebunk treatment | -1.26<br>(0.89)<br>β = -.03 | **-2.86****<br>**(0.91)**<br>β = -.07 | **-5.69****<br>**(2.06)**<br>β = -.11 | -1.08<br>(0.89)<br>β = -.02 | **-2.75****<br>**(0.91)**<br>β = -.07 | **-4.25***<br>**(1.73)**<br>β = -.08 |
| Debunk treatment | **-2.21***<br>**(0.89)**<br>β = -.05 | **-2.00***<br>**(0.91)**<br>β = -.05 | **-7.06*****<br>**(2.02)**<br>β = -.14 | **-2.17***<br>**(0.89)**<br>β = -.05 | **-1.89***<br>**(0.91)**<br>β = -.05 | **-5.19****<br>**(1.70)**<br>β = -.10 |
| Disinformation included in M1 | | | 1.60<br>(2.01)<br>β = .04 | | | 0.90<br>(1.70)<br>β = .02 |
| Controls included | No | No | No | Yes | Yes | Yes |
| Constant | 4.91***<br>(1.08) | 8.10***<br>(0.96) | 41.35***<br>(1.45) | 4.49*<br>(1.92) | 9.79***<br>(1.98) | 37.94***<br>(3.29) |
| N | 925 | 925 | 925 | 925 | 925 | 925 |
| Adjusted $R^2$ | 0.66 | 0.58 | 0.02 | 0.66 | 0.59 | 0.31 |
| | Wald test of the difference between Prebunk and Debunk | | | | | |
| $\chi^2(1)$ | 0.86 | 0.66 | 0.23 | 1.13 | 0.66 | 0.15 |
| p-value | 0.36 | 0.42 | 0.63 | 0.29 | 0.42 | 0.70 |

*Note.* The table reports OLS regressions predicting trust in true, false, and disinformation statements in M2, without (Columns 1–3) and with (Columns 4–6) control variables. Categorical variables are dummy-coded. For each predictor, unstandardized coefficients, standard errors (in parentheses), and standardized coefficients (β) are shown. Wald tests comparing Prebunk and Debunk treatments appear at the bottom. Statistically significant predictors ($p < .05$) are in bold. $p < .05$; $p < .01$; $p < .001$.

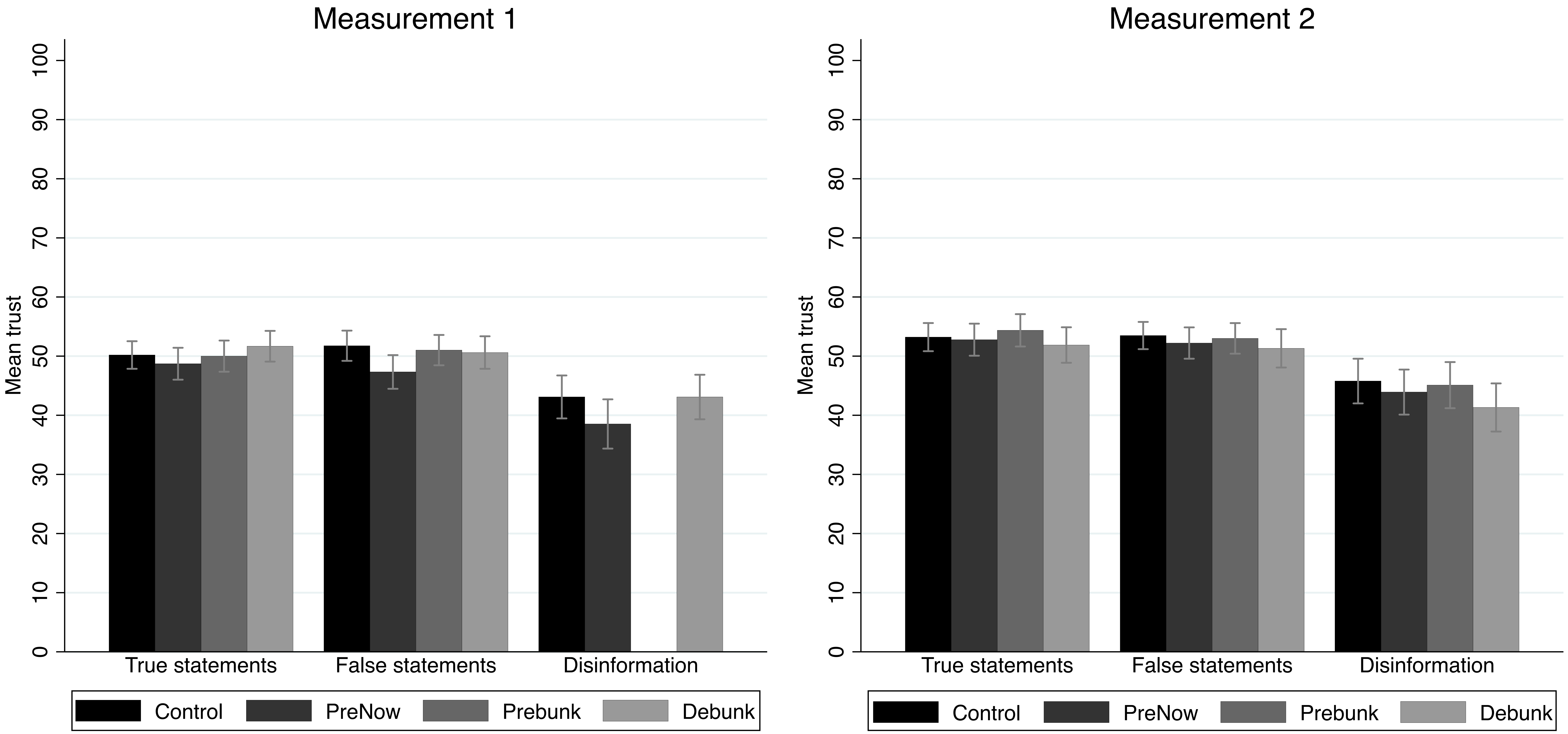
# Online experiment 2

- Representative sample, online, but:
  - Measurements 2 weeks apart
  - Participant forced to stay on a screen with intervention text for some time, text presented line by line
  - New **PreNow treatment**

# Experimental design

| | Measurement 1 (Pretest) | | | Measurement 2 (Posttest) | |
|---|---|---|---|---|---|
| **Treatment** | **Statements** | **Information** | **Other** | **Statements** | **Other measures** |
| **Control** | 10 true<br>5 false<br>5 disinfo | Neutral message about war after statements | Attitudes towards Russo-Ukrainian war<br>Blame for war<br>Reading literacy<br>Demographics | 10 true<br>5 false<br>5 disinfo | Exposure to fake news outside of the experiment<br>Political orientation<br>Feelings of threat<br>Status anxiety<br>Political cynicism<br>Cognitive reflection |
| **PreNow** | 10 true<br>5 false<br>5 disinfo | Disinformation refuting message before statements | | 10 true<br>5 false<br>5 disinfo | |
| **Prebunk** | 10 true<br>5 false | Disinformation refuting message after statements | | 10 true<br>5 false<br>5 disinfo | |
| **Debunk** | 10 true<br>5 false<br>5 disinfo | Disinformation refuting message after statements | | 10 true<br>5 false<br>5 disinfo | |

# Online experiment 2 results



# Online experiment 2

**Table 5: OLS regressions, Experiment 3 (abbreviated)**

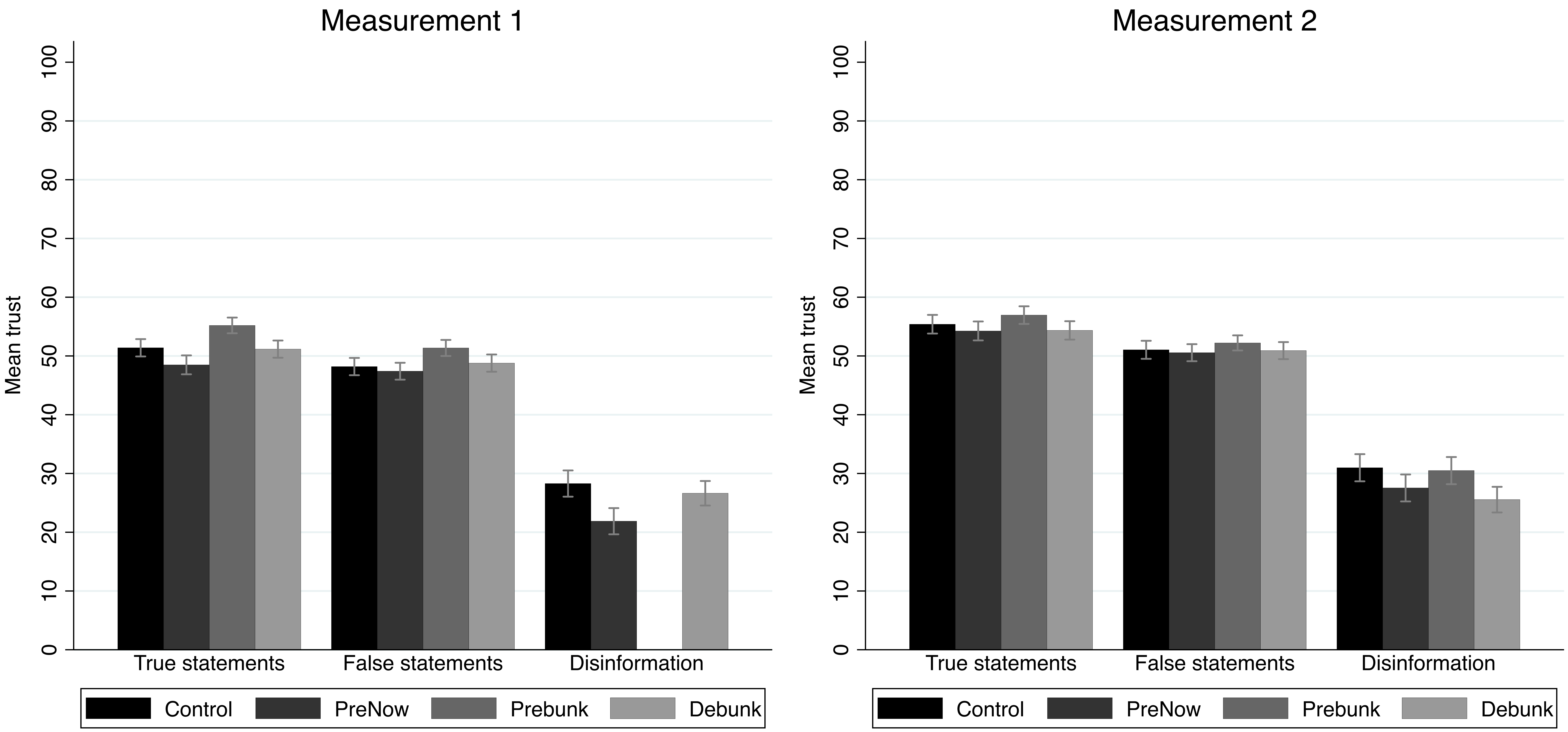
| | (1) True | (2) False | (3) Disinfo | (4) True | (5) False | (6) Disinfo |
|---|---|---|---|---|---|---|
| Corresponding M1 trust | **0.70*****<br>**(0.03)**<br>β = .66 | **0.51*****<br>**(0.03)**<br>β = .51 | | **0.65*****<br>**(0.03)**<br>β = .61 | **0.49*****<br>**(0.03)**<br>β =.49 | |
| PreNow treatment | 0.59<br>(1.49)<br>β = .01 | 1.00<br>(1.71)<br>β = .02 | -1.86<br>(2.82)<br>β = -.03 | 0.18<br>(1.48)<br>β = .00 | 0.82<br>(1.71)<br>β = .02 | **-4.51***<br>**(2.15)**<br>β = -.08 |
| Prebunk treatment | 1.28<br>(1.48)<br>β = .03 | -0.10<br>(1.69)<br>β = -.00 | -0.69<br>(2.80)<br>β = -.01 | 0.89<br>(1.47)<br>β = .02 | -0.49<br>(1.69)<br>β = -.01 | -2.20<br>(2.14)<br>β = -.04 |
| Debunk treatment | -2.37<br>(1.48)<br>β = -.06 | -1.58<br>(1.70)<br>β = -.04 | -4.47<br>(2.81)<br>β = -.08 | -2.47<br>(1.48)<br>β = -.06 | -1.52<br>(1.70)<br>β = -.04 | **-6.14****<br>**(2.15)**<br>β = -.11 |
| Controls included | No | No | No | Yes | Yes | Yes |
| Constant | 18.32***<br>(1.89) | 26.95***<br>(2.15) | 45.79***<br>(2.00) | 18.83***<br>(2.59) | 22.88***<br>(2.92) | 43.96***<br>(3.18) |
| N | 655 | 655 | 655 | 655 | 655 | 655 |
| Adjusted $R^2$ | 0.43 | 0.25 | 0.00 | 0.44 | 0.26 | 0.42 |
| | Wald test of differences between treatments - $\chi^2(1)$ [p-value] | | | | | |
| PreNow vs. Prebunk | 0.22 [.64] | 0.42 [.52] | 0.17 [.68] | 0.23 [.63] | 0.60 [.44] | 1.16 [.28] |
| PreNow vs. Debunk | 3.99 [.05] | 2.30 [.13] | 0.86 [.35] | 3.23 [.07] | 1.91 [.17] | 0.58 [.45] |
| Prebunk vs. Debunk | 6.15 [.01] | 0.77 [.38] | 1.83 [.18] | 5.23 [.02] | 0.38 [.54] | 3.36 [.07] |

*Note.* The table reports OLS regressions predicting trust in true, false, and disinformation statements in M2, without (Columns 1–3) and with (Columns 4–6) control variables. Categorical variables are dummy-coded. For each predictor, unstandardized coefficients, standard errors (in parentheses), and standardized coefficients (β) are shown. Pairwise comparison of treatments (Wald tests) appears at the bottom. Full results are reported in the Supplementary Materials. Statistically significant predictors ($p < .05$) are in bold.
$p < .05$; p < .01; $p < .001$.

# Online experiment 3

- Representative sample, online, Measurements 2 weeks apart, PreNow treatment, but:
  - 3 domains of disinformation: Politics, Climate, Health

# Online experiment 3 results



# Online experiment 3

**Table 6: OLS regressions, Experiment 4 (abbreviated)**

| | (1)<br>True | (2)<br>False | (3)<br>Disinfo | (4)<br>True | (5)<br>False | (6)<br>Disinfo |
|---|---|---|---|---|---|---|
| Corresponding M1 trust | **0.70*****<br>**(0.02)**<br>β = .67 | **0.57*****<br>**(0.02)**<br>β = .58 | | **0.63*****<br>**(0.02)**<br>β = .60 | **0.56*****<br>**(0.02)**<br>β = .57 | |
| PreNow treatment | 0.89<br>(0.85)<br>β = .03 | -0.03<br>(0.84)<br>β = -.00 | **-3.44***<br>**(1.64)**<br>β = -.07 | 0.74<br>(0.82)<br>β = .02 | -0.04<br>(0.84)<br>β = -.00 | -2.50<br>(1.36)<br>β = -.05 |
| Prebunk treatment | -1.11<br>(0.86)<br>β = -.03 | -0.65<br>(0.86)<br>β = -.02 | -0.49<br>(1.66)<br>β = -.01 | -0.94<br>(0.83)<br>β = -.03 | -0.58<br>(0.85)<br>β = -.02 | -0.21<br>(1.37)<br>β = -.00 |
| Debunk treatment | -0.89<br>(0.85)<br>β = -.03 | -0.47<br>(0.85)<br>β = -.02 | **-5.43*****<br>**(1.64)**<br>β = -.11 | -0.74<br>(0.82)<br>β = -.02 | -0.46<br>(0.84)<br>β = -.02 | **-4.11****<br>**(1.37)**<br>β = -.09 |
| Controls included | No | No | No | Yes | Yes | Yes |
| Constant | 19.41***<br>(1.30) | 23.41***<br>(1.26) | 30.98***<br>(1.17) | 3.53<br>(2.86) | 12.07***<br>(3.05) | -7.84<br>(4.65) |
| N | 1266 | 1266 | 1266 | 1266 | 1266 | 1266 |
| Adjusted $R^2$ | 0.44 | 0.33 | 0.01 | 0.48 | 0.34 | 0.32 |
| | Wald test of differences between treatments - $\chi^2(1)$ [p-value] | | | | | |
| PreNow vs. Prebunk | 5.37 [.02] | 0.52 [.47] | 3.24 [.07] | 4.12 [.04] | 0.40 [.53] | 2.81 [.09] |
| PreNow vs. Debunk | 4.43 [.04] | 0.27 [.60] | 1.49 [.22] | 3.32 [.07] | 0.25 [.62] | 1.42 [.23] |
| Prebunk vs. Debunk | 0.07 [.80] | 0.04 [.84] | 8.97 [.00] | 0.06 [.81] | 0.02 [.89] | 8.07 [.00] |

*Note.* The table reports OLS regressions predicting trust in true, false, and disinformation statements in M2, without (Columns 1–3) and with (Columns 4–6) control variables. Categorical variables are dummy-coded. For each predictor, unstandardized coefficients, standard errors (in parentheses), and standardized coefficients (β) are shown. Pairwise comparison of treatments (Wald tests) appears at the bottom. Full results are reported in the Supplementary Materials. Statistically significant predictors ($p < .05$) are in bold.
$p < .05$; $p < .01$; $p < .001$.

**Table 7: OLS regressions - disinformation across domains, Experiment 4 (abbreviated)**

| | (1) Politics | (2) Climate | (3) Health | (4) Politics | (5) Climate | (6) Health |
|---|---|---|---|---|---|---|
| PreNow treatment | -3.01 | -0.73 | **-6.59*** ** | -0.84 | 0.20 | **-5.93*** ** |
| | (1.80) | (1.84) | **(1.95)** | (1.42) | (1.47) | **(1.53)** |
| | β = -.06 | β = -.01 | β = -.12 | β = -.02 | β = .00 | β = -.10 |
| Prebunk treatment | -0.51 | 0.72 | -1.67 | 0.11 | 1.81 | -2.53 |
| | (1.82) | (1.87) | (1.97) | (1.43) | (1.49) | (1.54) |
| | β = -.01 | β = .01 | β = -.03 | β = .00 | β = .03 | β = -.04 |
| Debunk treatment | **-3.74*** | -3.27 | **-9.27*** ** | -1.90 | -2.25 | **-8.27*** ** |
| | **(1.81)** | (1.85) | **(1.96)** | (1.42) | (1.48) | **(1.53)** |
| | β = -.07 | β = -.06 | β = -.16 | β = -.04 | β = -.04 | β = -.14 |
| Prior beliefs | | | | **5.37*** ** | **5.28*** ** | **5.36*** ** |
| | | | | **(0.29)** | **(0.40)** | **(0.34)** |
| | | | | β = .47 | β = .40 | β = .38 |
| Other controls | No | No | No | Yes | Yes | Yes |
| Constant | 28.92*** | 26.28*** | 37.73*** | -6.35 | -10.06 | -17.41** |
| | (1.28) | (1.32) | (1.39) | (5.00) | (5.56) | (5.74) |
| N | 1266 | 1266 | 1266 | 1266 | 1266 | 1266 |
| Adjusted $R^2$ | 0.00 | 0.00 | 0.02 | 0.38 | 0.37 | 0.41 |
| | Wald test of differences between treatments - $\chi^2(1)$ [p-value] | | | | | |
| PreNow vs. Prebunk | 1.92 [.17] | 0.61 [.44] | 6.31 [.01] | 0.45 [.50] | 1.18 [.28] | 4.94 [.03] |
| PreNow vs. Debunk | 0.17 [.68] | 1.92 [.17] | 1.91 [.17] | 0.57 [.45] | 2.79 [.10] | 2.38 [.12] |
| Prebunk vs. Debunk | 3.16 [.08] | 4.60 [.03] | 14.9 [.00] | 1.99 [.16] | 7.45 [.01] | 14.0 [.00] |

*Note.* The table reports OLS regressions predicting trust in disinformation statements in M2 in the domains of politics, climate and health, without (Columns 1–3) and with (Columns 4–6) control variables. Categorical variables are dummy-coded. For each predictor, unstandardized coefficients, standard errors (in parentheses), and standardized coefficients (β) are shown. Pairwise comparison of treatments (Wald tests) appears at the bottom. Full results are reported in the Supplementary Materials. Statistically significant predictors ($p < .05$) are in bold. $p < .05$; p < .01; $p < .001$.

# Summary

- Debunk intervention consistently and significantly reduces the trust in disinformation.
- Debunk yields similarly consistent trust-reducing effect also on true statements.
- Prebunk intervention seems to effectively lower trust in disinformation only
  - for a very short period or
  - if implemented moments before exposure to fake news.