Web-based experiment control software for research and teaching on human learning

Helena Matute, Miguel A. Vadillo, and Raúl Bárcena
Universidad de Deusto, Bilbao, Spain

In this article we describe some of the experimental software we have developed for the study of associative human learning and memory. All these programs have the appearance of very simple video games. Some of them use the participants' behavioral responses to certain stimuli during the game as a dependent variable for measuring their learning of the target cue-outcome associations. Some others explicitly ask participants to rate the degree of relationship they perceive between the cues and the outcomes. These programs are implemented in Web pages using JavaScript, which allows their use both in traditional laboratory experiments as well as in Internet-based experiments.

The psychology of learning is a research area that has usually been investigated with nonhuman animals and in which, traditionally, there existed too many procedural and ethical problems to conduct experiments with humans. However, human learning is today a flourishing research area in which many interesting effects are being reported around the world (see, e.g., De Houwer & Beckers, 2002; Shanks, 2007; Shanks, Holyoak, & Medin, 1996, for reviews).

But the main problem when a researcher decides to run an experiment on human learning is the lack of commercial equipment. In principle, a personal computer is all that is needed, but the experiment needs to be programmed. All researchers working with human subjects have necessarily started their research by creating their own experimental preparations: software, instructions, measurements, and techniques. This is a costly process that takes a very long time and a great effort. However, it could be easily simplified if these programs were shared among researchers. With this in mind we here present a brief description of some of the programs that our research group has developed over the past 10 or 15 years. Demonstrations of these programs can be freely downloaded from our Web site: www.labpsico.com. They can be used as Internet-based experiments or as laboratory experiments in Web-based courses. In the later case they can be used the way they are presented. In the first case they need to be adapted to the researchers' needs by changing the design of the experiment in the source code: Number of trials, duration of stimuli, number of groups, and so on.

All the experimental programs we present here have been designed to study how associations between events are acquired, retrieved and used. Some of the programs assess the learning of stimulus–stimulus (S–S) associations (i.e., akin to Pavlovian conditioning), while the other ones assess the learning of response–outcome (R–O) associations (i.e., akin to instrumental learning). The main difference with respect to animal conditioning is that we are not using biologically significant events as unconditioned stimuli (US), and therefore we cannot assess conditioned responses. But there are other indexes of the acquisition of an association between two events. Some of the software, the so-called judgmental tasks, assess participants' learning of the associations by directly asking participants to rate the perceived strength of the relationship between the two events. Alternatively, the so-called behavioral tasks assess participants' learning of the associations by measuring the participants' behavioral responses to the stimuli being presented during the experiment.

All of these programs have the appearance of very simple computer games, which facilitates the recruitment of voluntary research participants. Even though these games are by far simpler than those that many students are used to playing, students generally find this format of presentation of experiments interesting. Moreover, this format also allows researchers to use symbolic reinforcement (e.g., points in a game) and punishment (e.g., a Martians' invasion) so as to avoid some of the traditional problems associated to the use of true USs in human research.

In general, what can be observed in these experiments is that during the first trials subjects do not know what to say or do; however, as learning proceeds a learning curve can be observed and this shows either that the behavioral responses to stimuli increase—or decrease—or that the subjective judgment about the relationship increases—or decreases—as a result of training. These trial-by-trial results are stored in a .txt data file that is created when the experiment is run and that can be easily exported to any statistical program for analysis. In this .txt file, if several participants are run in the same computer, the data from

---

H. Matute, matute@fice.deusto.es

each of them is added to that of the previous participant in the same computer. Each row shows the results of each subject, each column shows a different variable. In order to facilitate its use and subsequent modification, the data file that is created after running the demonstration file that we provide for each of these programs includes a first row in which we mention the names of the different variables that are being stored in the demonstration, and the order in which they get stored (e.g., name of the experiment, group, counterbalancing, number of responses in each training trial, number of responses in the test trial).

Most of these programs were first described in the methodological section of some of our research articles as being implemented in Spanish and in either QuickBasic or Borland C++ programs. Their current implementation on both Spanish and English Web pages using JavaScript allows a more general and flexible use. Specifically, these tasks can be used to run experiments not only in the traditional laboratory conditions, but also over the Internet with anonymous participants from all over the world. Most importantly, they allow researchers to run an experiment simultaneously both in the laboratory and over the Internet, a strategy that is proving to be very useful in order to assess the generality of laboratory-based experiments (Vadillo, Bárcena, & Matute, 2006; Vadillo & Matute, 2007; Vadillo, Miller, & Matute, 2005).

Below, we briefly describe some of these programs with reference to some of the learning phenomena that we have demonstrated with each of them. Note however that each of them can be adapted to implement very different experiments from the ones we mention here. For convenience, the source code is provided with the downloadable files. It is relatively easy for anyone with some knowledge of JavaScript (or of computer programming, in general) to adapt these programs to different experimental designs and specifications: JavaScript code is inserted in htm files, and thus, it can be edited using any text editor or Web design software. Changing the number of trials, for example, is as simple as adding or deleting lines in the trial sequence, which can easily be recognized even if one knows nothing about JavaScript (but knows something about programming). Changing other things, like the particular instructions that the participants receive, the wording of the questions they receive at test, or the particular stimuli (pictures) they see during the experiment is achieved by just changing the text of the instructions or questions or changing the picture that the program is showing during each trial, in a way similar to what one would do to modify text or pictures in a Web page. Perhaps, some other changes, such as adding completely new groups or counterbalancing conditions to the study, could be more difficult. But if one finds problems here, it is always possible to use tricks such as, for example, copying the same experiment several times and running a different group or counterbalancing condition in each version of the program (such as what is done, for example, in commercial software such as SuperLab).

STIMULUS–STIMULUS TASKS

Martians

This is a behavioral task designed as an analogue of the conditioned suppression technique that is commonly used in animal learning research. In animal research, when a cue (e.g., a light) is paired to an electric shock, the animal learns to predict shock when the light is presented. As a consequence, if the animal is pressing a bar to obtain food and the light is presented, the animal will freeze. The strength of this freezing behavior is assessed as the degree in which bar pressing behavior is suppressed. Thus, suppression of bar pressing is used as a dependent variable to assess whether the animal has learned the predictive relationship between the light and the shock.

Our Martians task also uses suppression of bar pressing as an index that humans have learned the predictive relationship that exists between two events. In this case, participants first learn to press the space bar of a computer (i.e., their "laser gun") regularly in order to prevent Martians from landing on the earth. Once participants learn this, an instructional screen tells them that the Martians have developed an anti-laser shield and that if they shoot (i.e., if they press the bar) while the shield is connected, their shot will be reflected back to them and thousands of Martians will immediately invade. Participants are also told that they will know that the shield is connected when they see a black and white flash, and that, therefore, they should refrain from responding when they see this black and white flashing screen. Thus, the flashing screen is an analogue to the electric shock producing freezing in animal research, only that the flash is here given its aversive motivational value through instructions. The interesting point is that during the game, we usually present some predictive cues (e.g., tones, colors, or other cues). Some of them predict that the flash will follow, others that it will not. Participants are never told to refrain from responding while these cues are present, nor is responding in their presence punished. However, participants show a gradual learning curve that reflects increasing suppression of responding to the cues that predict that the flash is going to occur. Conversely, they show no suppression to the cues that predict that the flash will not occur.

Using this procedure our research team has demonstrated acquisition and extinction (Arcediano, Ortega, & Matute, 1996), blocking (Arcediano, Matute, & Miller, 1997), retroactive interference (Matute & Pineño, 1998), instructional effects (Arcediano et al., 1996), and contextual modulation of responding (Matute & Pineño, 1998). Other research teams have also used this task to study effects such as blocking (Vandenbrouke, 1999), miscuing and retroactive interference (Lipp & Dal Santo, 2002), and extinction and contextual modulation (Havermans, Keuker, Lataster, & Jansen, 2005). In principle, any experiment that could use suppression ratio with animals as the index that the association has been acquired can also be conducted with humans using this task.

Spy Radio

This is another behavioral task that assesses S–S learning. In the above mentioned Martians task, learning is measured by looking at the disruption of the ongoing behavior elicited by the presentation of a predictive stimulus. An alternative way to assess the learning of an association between two events is to build a task in which the production of a response, rather than its suppression, is the dependent variable. Thus, an active response (pressing the space bar) is used in this program to assess the strength with which the participant expects the outcome to occur in each trial.

The instructions of the game tell participants that they are supposed to be drivers of a truck that they can use to take people out from a city that is in the middle of a war. The truck is equipped with a spy radio that can receive codified information about the state of the road, so that the driver can know whether the road is safe or mined. The interesting point is that the state of the road can be predicted before letting people get onto the truck by looking at the color lights appearing on the spy radio panel. Some colors indicate that the road is safe and others that the road is mined, but participants are not told about this; they have to learn it from experience. Thus, by paying attention to the panel and to the consequences of each trip, participants can learn to predict whether the people they will drive in the truck will survive or not. Participants are given one point for each person they take out of the city to a safe place and they lose one point for each person killed by a mine during the trip. Therefore, as participants learn the association between each color and its usual outcome, they will take more people in the truck when the presented colors indicate that the road is safe and fewer when the cues indicate that the road is mined. Hence, the total amount of refugees that enter the truck will reflect a learning curve and can be used as a measure of the participant's learning of the target association between each color and its outcome.

Using this task we have investigated phenomena related to overshadowing (Vadillo, et al., 2006), blocking (Escobar, Pineño, & Matute, 2002), retroactive interference (Ortega & Matute, 2000; Pineño & Matute, 2000), and proactive interference (Castro, Ortega, & Matute, 2002). Other research teams have also used this task to study reasoning processes (Dieussaert, Schaeken, & d'Ydewalle, 2002).

Allergies

This task was not originally developed by our research group. Indeed, different versions of it have been developed and are being used in virtually all laboratories studying causal and predictive learning around the world. It is so convenient and easy to use that it has become a standard preparation in the study of human learning, particularly in the study of causal and predictive learning.

The allergy task (also known with the more generic name of "judgmental task") is slightly different from the other ones we have described so far: instead of assessing behavioral responses to stimuli it assesses subjective judgments about the strength of the relationship between two events.

In the version that is available in our Web site participants have to imagine that they are allergists who are investigating the potential secondary effects of some recently developed drugs. In order to test their effects, a sample of fictitious patients is provided. These patients take different combinations of the drugs that are going to be tested. Some of those patients develop an allergic reaction to the drugs and some others do not. In each trial the participant sees first the medical record indicating the name of the drug the fictitious patient has taken; then the participant is asked to predict whether or not this patient will develop the allergic reaction; finally, the participant is told whether or not the patient developed the allergy. When participants face the first trials they can only guess whether the reaction will appear on each patient. But as training proceeds they can base their predictions on the previous drug–allergy pairings. Thus, the functioning of this task is analogous to that of the behavioral tasks we described above: Participants' predictions can be used as an index of the underlying learning of the relevant associations between each drug and the allergic reaction. The participants' judgment about the strength of the relationship between the two events is usually assessed using a numerical scale that the participant can click using the mouse.

Using this preparation we have investigated cue competition effects (Matute, Arcediano, & Miller, 1996), extinction and contextual modulation of responding (Vadillo, Vegas, & Matute, 2004) as well as several procedural factors that usually vary from one version of this program to the other and that can strongly influence the results, such as, for example, the way in which the test question is worded, the frequency with which judgments are required and the instructions that participants receive (Matute, Vegas, & De Marez, 2002; Vadillo & Matute, 2007; Vadillo, Miller, & Matute, 2005). Other researchers have also demonstrated a wide variety of critical learning effects using different versions of this task (e.g., Allan, Siegel, & Tangen, 2005; Buehner & Cheng, 1997; Catena, Maldonado, & Cándido, 1998; Cobos, López, Caño, Almaraz, & Shanks, 2002; Dickinson & Burke, 1996; Karazinov & Boakes, 2004; Le Pelley & McLaren, 2001; Vila & Rosas, 2001; Waldmann, 2001; Wasserman, 1990).

Folders and Cards

This judgmental task is a variation of the allergy task that we developed in order to study the directionality of associative learning. That is, when participants learn that Cue A is associated with Outcome 1, will they be able to infer the presence of Cue A upon detecting the presence of Outcome 1?

This task was developed in two different versions, one with folders, the other one with cards (see Gerolin & Matute, 1999). In the Cards version, the cover story introduced participants in a card game situation. Each card had two sides, one colored and the other one containing a figure. During each training trial the color side was always presented first. During the test, the figure side was presented and participants had to tell which color was on the other side. The dependent variables are (1) whether or not the participant gives the correct response and (2) the degree of certainty (in a numerical rating scale) with which that response is emitted.

The second version, the Folders version, was very similar to the Cards version but rather than cues and outcomes being two sides of a card, cues were folders of different colors and each of them contained a piece of paper showing a different figure (outcome). This is a simple variation that was aimed just to make it explicit that the color represented one stimulus (folder) and the figure a different one, so that the cue and the outcome could not be interpreted as a configural stimulus (as was the case in the Cards version). The implementation of this task that can be downloaded from our Web site shows this Folders version, but the only change that is needed in order to use the more configural Cards game is to use a picture of a card instead of a picture of a folder in the program, as well as a change in the instructions (see Gerolin & Matute, 1999 for the instructions used in both versions).

Using this task we have so far shown that associations can be used in the direction opposite to the one they were learned (Gerolin & Matute, 1999). Nevertheless, this task can potentially be used in a much wider range of associative phenomena. After all, this task is no different from the other ones described here in which different cues predict different outcomes. Therefore, all basic associative learning and memory phenomena should also be replicable using this task. In fact, other researchers have used very similar tasks to investigate other learning effects such as blocking (e.g., Glautier, 2002).

RESPONSE–OUTCOME TASKS

Flashes

All the tasks mentioned above were designed to study how people perceive that different events of their environment are related. The flashes task was developed to study the mechanisms by which people discover whether or not their own behavior (instead of an external event) causes or prevents a given outcome. In this task participants are told that their goal is to control the appearance of flashes in the computer's screen using the keyboard. Depending on the researcher's interests participants can be asked either to produce these flashes or to prevent their occurrence. Then, after being exposed to a given sequence of trials participants are asked to assess the degree to which they think they have been able to control the flashes.

This flashes task is an adaptation of a previous preparation developed by Matute (1996) for the study of the illusion of control (i.e., people's tendency to believe that they have control over uncontrollable outcomes under certain conditions; see Langer, 1975). The original preparation used tones rather than flashes; in the present version we are using flashes in order to facilitate the execution of these experiments over the Internet.

The experiments we are running using the flashes task are also dealing with the illusion of control. The sequence of flashes that appears on the screen follows a preprogrammed schedule, which means that participants cannot alter this sequence. However, when they are asked to rate their degree of control, the mean judgments of control they provide are generally significantly greater than zero.

That is, even though participants have no control over the flashes they still perceive that it is their behavior that is producing the sequence of flashes. This illusion of control can be observed even when participants are warned about the possibility that the flashes might be uncontrollable. Moreover, this effect is not only observed when the experiment is run in the laboratory, but also when it is performed over the Internet (Matute, Vadillo, Vegas, & Blanco, 2007).

Although the task was originally designed to study people's reactions when they are exposed to uncontrollable events, it can also be adapted to the study of people's instrumental behavior under situations in which they do have the opportunity to influence the probability of an outcome's occurrence.

CONCLUDING COMMENTS

A first step toward a time in which the different experimental procedures are sufficiently standardized so that anyone can run an experiment without needing to previously invent the procedure consists of sharing preparations among researchers. The only human preparation that has so far become a standard is the allergy (or "judgmental") task. As noted previously, there are so many researchers using so many variations of the allergy task that this is allowing the scientific community to develop an excellent data base of those variables that do affect the results in this task and how they do so. Our purpose has been to enlarge the horizon of tasks that can be shared and tested among researchers.

But why should we share and use other tasks in addition to the allergy task if that one has already become standard? We believe there are at least two reasons for doing so.

First, although the different tasks we have described here have been developed as general purpose preparations that in principle should be adaptable to almost any experimental design, it would not be surprising that some of them are better suited to study some effects than others. For example, experiments involving the acquisition of associations between causes and effects could probably best be run using the allergy task, whereas experiments involving configural stimuli could probably best be run using the Cards game, and experiments on predictive learning could use either the Martians or the spy radio preparations.

A second advantage of sharing a variety of preparations is that, in almost all universities, the average psychology student takes part in several psychology experiments per year… and different researchers are often not entirely aware of how the other experiments being simultaneously conducted on campus (or conducted in the same laboratory one or two years before) look like. The greater the number of different preparations available, the more different these experiments will look from one another, and thus, the smaller the chances that one of them will affect the results of the other one. Of course, the results will always need to be controlled through careful designs and replications, but using a variety of different preparations will certainly help get cleaner data. Moreover, if these experiments can—as

the ones we have shown here—be executed over the noisy environment of the Internet, then what we have is something that is at least as important as having a shared methodology: a methodology that allows one to test the generality of the results outside the psychology laboratory.

Of course, there are also some risks associated to the sharing of procedures that we propose. First, there could be an unknown confound in the procedure which could invalidate everything done with that procedure. Second, using a standardized procedure decreases the likelihood of a revolutionary new finding or perspective. However, there are different stages in the development of a science that result most benefited from efforts to unify procedures. We believe that research on human learning is still in need of some methodological consensus.

AUTHOR NOTE

REFERENCES

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior*, **33**, 250-263.

Arcediano, F., Matute, H., & Miller, R. R. (1997). Blocking of Pavlovian conditioning in humans. *Learning & Motivation*, **28**, 188-199.

Arcediano, F., Ortega, N., & Matute, H. (1996). A behavioral preparation for the study of human Pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, **48B**, 270-283.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory versus the Rescorla–Wagner model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Sciences Society* (pp. 55-60). Mahwah, NJ: Erlbaum.

Castro, L., Ortega, N., & Matute, H. (2002). Proactive interference in human predictive learning. *International Journal of Comparative Psychology*, **15**, 55-68.

Catena, A., Maldonado, A., & Cándido, A. (1998). The effect of the frequency of judgment and the type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 481-495.

Cobos, P. L., López, F. J., Caño, A., Almaraz, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, **28**, 331-346.

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology*, **55B**, 289-310.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, **49B**, 60-80.

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The quality of test context and contra-evidence as a moderating factor in the belief revision process. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 280-285). Mahwah, NJ: Erlbaum.

Escobar, M., Pineño, O., & Matute, H. (2002). A comparison between elemental and compound training of cues in retrospective revaluation. *Animal Learning & Behavior*, **30**, 228-238.

Gerolin, M., & Matute, H. (1999). Bidirectional associations. *Animal Learning & Behavior*, **27**, 42-49.

Glautier, S. (2002). Spatial separation of target and competitor cues enhances blocking of human causality judgements. *Quarterly Journal of Experimental Psychology*, **55B**, 121-135.

Havermans, R. C., Keuker, J., Lataster, T., & Jansen, A. (2005). Contextual control of extinguished conditioned performance in humans. *Learning & Motivation*, **36**, 1-19.

Karazinov, D. M., & Boakes, R. A. (2004). Learning about cues that prevent an outcome: Conditioned inhibition and differential inhibition in human predictive learning. *Quarterly Journal of Experimental Psychology*, **57B**, 153-178.

Langer, E. J. (1975). The illusion of control. *Journal of Personality & Social Psychology*, **32**, 311-328.

Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *Quarterly Journal of Experimental Psychology*, **54B**, 311-352.

Lipp, O. V., & Dal Santo, L. (2002). Cue competition between elementary trained stimuli: US miscuing, interference, and US omission. *Learning & Motivation*, **33**, 327-346.

Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychological Science*, **7**, 289-293.

Matute, H., Arcediano, F., & Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 182-196.

Matute, H., & Pineño, O. (1998). Stimulus competition in the absence of compound conditioning. *Animal Learning & Behavior*, **26**, 3-14.

Matute, H., Vadillo, M. A., Vegas, S., & Blanco, F. (2007). The illusion of control in Internet users and college students. *CyberPsychology & Behavior*, **10**, 176-181.

Matute, H., Vegas, S., & De Marez, P. J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 714-725.

Ortega, N., & Matute, H. (2000). Interference between elementally trained stimuli can take place in one trial. *Learning & Motivation*, **31**, 323-344.

Pineño, O., & Matute, H. (2000). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology*, **13**, 16-33.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, **60**, 291-309.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.) (1996). *The psychology of learning and motivation: Vol. 34. Causal judgment.* San Diego: Academic Press.

Vadillo, M. A., Bárcena, R., & Matute, H. (2006). The Internet as a research tool in the study of associative learning: An example from overshadowing. *Behavioural Processes*, **73**, 36-40.

Vadillo, M. A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology*, **60**, 433-447.

Vadillo, M. A., Miller, R. R., & Matute, H. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior*, **33**, 172-183.

Vadillo, M. A., Vegas, S., & Matute, H. (2004). The frequency of judgment as a context-like determinant of predictive judgments. *Memory & Cognition*, **32**, 1065-1075.

Vandenbroucke, A. (1999, May). *Blocking in the Martians preparation.* Paper presented at the Cue Competition Meeting, Lignely, Belgium.

Vila, N. J., & Rosas, J. M. (2001). Reinstatement of acquisition performance by the presentation of the outcome after extinction in causality judgment. *Behavioural Processes*, **56**, 147-154.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, **8**, 600-608.

Wasserman, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science*, **1**, 298-302.