Supplement S1

Transparent Reporting and Explanation of Deviations from the Preregistrations

This supplement presents the preregistrations for Experiments 2 and 3. Experiment 1 and the pilot study to Experiment 3 were not preregistered.

Location of Preregistrations

The preregistrations can be found on aspredicted.org under the following links.

Exp. 2: https://aspredicted.org/7G9_D7Q

Exp. 3: https://aspredicted.org/BJH_F6F

Preregistered Hypotheses

In the preregistrations, predictions are clustered by measure (i.e., evaluative ratings, memory measures, MPT model). For increased clarity, preregistered hypotheses are clustered in this supplement by basic memory effects, efficiency of learning with anterior and posterior validity information, and observable consequences of the different controlled processes supposedly induced by the two learning procedures. The main article focuses on the latter. The preregistrations of the two studies are identical. The data files and analysis code are available on the OSF (https://osf.io/hfv9g/?view_only=c1d789fa21cd4f25bca0547f3b18395e).

Basic Memory Effects

(1) "Over the time interval, standard and reversed EC effects will become smaller."

This prediction was derived from the general memory literature that established forgetting as a function of time passing (Ebbinghaus, 1885/1913) and/or retroactive interference (Barnes & Underwood, 1959). Both effects have been demonstrated for evaluative conditioning (Förderer & Unkelbach, 2012; Gast et al., 2012; Sweldens et al., 2010; but see Hütter et al., 2012). This prediction was generally confirmed, although no significant changes were observed in the invalid condition of Experiment 3 with posterior validity information (applying one-tailed tests).

The prediction is not discussed in detail in the main article, but the result is presented.

Efficiency of Learning with Anterior vs. Posterior Validity Information

(2) "Standard and reversed EC effects should be larger in the advanced-instructions condition than in the postponed-instructions condition."

This hypothesis was formulated in anticipation of differences in the efficiency of learning in the two versions of the paradigm. As the posterior condition separates the US valence from the validity information in time, the joint encoding of CS, US valence, and validity information should be impaired. Additionally, the integration and interpretation at the judgment stage might take longer in the posterior condition. Due to this ambiguity, this hypothesis is not discussed in the main article, although the prediction was confirmed in both experiments as reported in the results sections.

(3) "The c-parameter will be smaller in the postponed-instructions condition than in the advanced-instructions condition (both at Time 1 and Time 2)."

The *c*-parameter reflects an evaluative judgment in line with the stated validity of a pairing. Therefore, the same prediction holds as for the evaluative ratings. As is reported in the main article, this hypothesis was generally confirmed with the exception that the difference between *c*-parameters did not reach significance at Time 2 in Experiment 3.

(4) "The effect of the time interval on evaluative ratings should be stronger in the postponed-instructions condition than in the advanced-instructions condition."

Given the generally low forgetting rates obtained on the evaluative ratings, it may not be surprising that this prediction was generally not confirmed. In Experiment 3, we obtained an effect opposite to the hypothesis in the valid condition ($F[1,158] = 7.29$, $p = .008$, $\eta_p^2 = .044$). This prediction and the according results are not presented in the main article.

(5) "Memory will be less accurate in the postponed-instructions condition than in the advanced-instructions condition."

This prediction is based on the fact that the joint encoding of US valence and US validity is more difficult in the posterior than the anterior condition. This general hypothesis was not confirmed, however. Overall (i.e., in an analysis that combines valence and validity memory across the first and second experimental session), there was no difference between the anterior and posterior conditions (Exp. 2: $F(1,158) = 0.38$, $p = .539$, $\eta_p^2 = .002$; Exp. 3: $F(1,158) = 0.34$, $p = .561$, $\eta_p^2 = .002$). This finding is encouraging as it allows interpreting differences between the timing of validity information condition in valence vs. validity memory (see main article for the report of the results of analyses comparing valence and validity memory).

(6) "Memory for instructions will decay more strongly in the postponed-instructions condition than in the advanced-instructions condition."

Validity memory does not serve as an unequivocal indicator of controlled processes as validity information might not be retained in both conditions, albeit for different reasons. In the anterior condition, validity information may be forgotten, because it is not needed if the controlled process performed an integration. In the posterior condition, it may be forgotten, because the representations may not be as stable. The planned analyses thus merely target potential difficulties at encoding influencing memory accuracy (over time), but are not informative on the nature of the controlled process.

In Experiment 2, validity memory was worse in the posterior than the anterior condition, $F(1,158) = 3.64$, $p = .058$, $\eta_p^2 = .023$, although this effect did not reach significance. However, when assessing the trajectory of memory accuracy from T1 to T2, the time of measurement × timing of validity interaction was not significant in Experiment 2, $F(1,158) = 0.005$, $p = .945$, $\eta_p^2 = .000$. Thus, even though there may be differences at encoding, validity information does not fade faster in the posterior than the anterior condition once encoding was successful.

In Experiment 3, validity memory was overall not worse in the posterior compared to the anterior condition, $F(1,158) = 0.00$, $p = .971$, $\eta_p^2 = .000$. However, the effect of time on memory was larger in the anterior condition, $F(1,158) = 9.83$, $p = .002$, $\eta_p^2 = .059$. Hence, with a relational encoding scheme, we obtained initial evidence that validity information is encoded equally well independent of the timing of validity information. However, validity information is less well retained when it was provided before rather than after the valence information.

Against our expectation, the posterior condition does not generally show worse memory or stronger decay. Together, the assessment of the predictions in this section show that the posterior condition does not generally hinder learning, but that the timing of validity information influences the emphasis of valence versus validity information. As shown by the differences in the EC effects and *c*-parameters, this differential learning has consequences for participants' evaluations of the CSs.

(7) "Meaning will be more successfully preserved in the advanced-instructions condition."

This prediction reflects the assumed role of meaning reversals due to lost validity tags. Forgetting of validity information may be less consequential – and admissible - when integration has taken place. We thus analyzed the prevalence of fully correct (i.e., both valence and validity memory are correct) and fully incorrect (i.e., both valence and validity memory are incorrect) responses, and checked for a main effect of timing of validity information.

In Experiment 2, we obtained a small effect in the predicted direction, which was not significant overall, $F(1,158) = 0.85$, $p = .359$, $\eta_p^2 = .005$, and did not depend on time of measurement, $F(1,158) = 1.11$, $p = .293$, $\eta_p^2 = .007$ (Exp. 3: $F(1,158) = 6.85$, $p = .010$, $\eta_p^2 = .042$). Observing the predicted effect would likely have required a larger time lag between learning and measurement (i.e., several days instead of 24 hours).

In Experiment 3, the predicted effect was significant, $F(1,158) = 7.15$, $p = .008$, $\eta_p^2 = .043$) and depended on time of measurement, $F(1,158) = 6.85$, $p = .010$, $\eta_p^2 = .042$. The effect was smaller at T2, $F(1,158) = 3.32$, $p = .070$, $\eta_p^2 = .021$, than at T1, $F(1,158) = 10.36$, $p = .002$, $\eta_p^2 = .062$ (possibly due to the larger forgetting rates for the validity information in the anterior condition; see point (6) above).

(8) "Over time the c-parameter will more strongly decrease in the postponed-instructions condition than in the advanced-instructions condition."

In Experiment 2 we obtained a significant decrease in the *c*-parameter over time, $F(1, 158) = 81.22$, $p < .001$, $\eta_p^2 = .340$, that did not depend on the timing of validity information, $F(1, 158) = 0.15$, $p = .699$, $\eta_p^2 = .001$.

Experiment 3 shows a significant decrease in the *c*-parameter over time, $F(1, 158) = 25.56$, $p < .001$, $\eta_p^2 = .139$, that depends on the timing of validity information, $F(1, 158) = 3.93$, $p = .049$, $\eta_p^2 = .024$. Against expectations, this latter effect indicates a slightly stronger decrease in the anterior validity information condition.

Together, these findings suggest that the posterior condition does not induce higher forgetting rates. In other words, to the degree that we obtain evidence speaking to separate encoding of US valence and US validity, they do not deteriorate faster than with integration.

Consequences of Different Controlled Processes

Erroneous (Valence) Memory

(9) "Valence will be more successfully preserved in the postponed-instructions condition."

In an analysis that regards valence memory performance from both sessions (i.e., T1 and T2), in Experiment 2, memory was indeed better in the posterior condition than the anterior condition (Exp. 2: $F[1,158] = 8.15$, $p = .005$, $\eta_p^2 = .050$). The severity of forgetting, however, did not differ between the timing conditions, (Exp. 2: $F[1,158] = 0.17$, $p = .684$, $\eta_p^2 = .001$). Thus, in line with our prediction valence is indeed preserved better under conditions assumed to foster separate encoding. However, we do not obtain differential effects of forgetting once encoding was successful.

In Experiment 3, valence memory did not differ significantly between timing conditions, $F(1,158) = 0.87$, $p = .354$, $\eta_p^2 = .005$. As in Experiment 2, the decrease in memory accuracy between measurement points did also not depend on timing of validity information, $F(1,158) = 1.58$, $p = .210$, $\eta_p^2 = .010$.

As a main consequence of the assumed controlled processes, this prediction is introduced and tested comprehensively in the main article.

(10) "False memories will indicate fusion representations in the advanced-instructions condition more so than in the postponed-instructions condition: If fusion models underlie the exertion of control then preserved meaning should be more often affirmed under reversal instructions as compared to preserved meaning being negated in apply condition (baseline error)."

This prediction is based on the assumption that integration leads to a higher prevalence of double errors in the condition that requires a reversal compared to the valid condition. That is, if a positive pairing is termed invalid and participants successfully form a CS-negative representation, they should be more inclined to indicate that the pairing was negative and that it had to be applied (cf. Mayo et al., 2004). Double errors in the standard condition serve as control for memory performance.

However, since pre-registering the studies, we came to prescind from this analysis as a particularly informative consequence of integration. The reason is that it might not be more informative than the analysis of the accuracy of valence memory alone. That is, if the prevalence of one type of error increases (i.e., valence), even if this occurs independent of the errors for validity information, the prevalence of double errors will also naturally increase. Hence, while a higher frequency of double errors is indeed observed in the anterior compared to the posterior condition, this finding may not give us additional information over and above the finding on valence errors. Thus, we focus on valence errors in the main body of the paper.

Nevertheless, to test the present hypothesis, we analyzed the absolute frequencies of double errors in a 2 (validity information) × 2 (time of measurement) × 2 (timing of validity information) ANOVA. In Experiment 2, the effect of validity information (i.e., the difference between the valid and invalid conditions) depended on the timing of the validity information, $F(1, 158) = 9.73$, $p = .002$, $\eta_p^2 = .058$. This effect was independent of time of measurement, $F(1, 158) = 1.36$, $p = .246$, $\eta_p^2 = .009$. Separate analyses per timing condition revealed that the main effect of validity information was significant in the anterior validity information condition, $F(1, 79) = 7.67$, $p = .007$, $\eta_p^2 = .088$, but not in the posterior validity information condition, $F(1, 79) = 2.08$, $p = .153$, $\eta_p^2 = .026$.

In Experiment 3, the effect of validity information (i.e., the difference between the valid and invalid conditions) depended on the timing of the validity information, $F(1, 158) = 15.51$, $p < .001$, $\eta_p^2 = .089$. This effect was independent of time of measurement, $F(1, 158) = 0.42$, $p = .520$, $\eta_p^2 = .003$. Separate analyses per timing condition revealed that the main effect of validity information was significant in the anterior validity information condition, $F(1, 79) = 24.12$, $p < .001$, $\eta_p^2 = .234$, but not in the posterior validity information condition, $F(1, 79) = 0.02$, $p = .894$, $\eta_p^2 = .000$.

Valence Asymmetries

(11) "An instructions x US valence interaction [on valence memory performance] indicating fusion representations will be stronger in the advanced-instructions condition than in the postponed-instructions condition."

Based on the often documented higher diagnosticity (Skowronski & Carlston, 1989) and resistance to change (Baumeister et al., 2001) of negative information, it was hypothesized that it may be more difficult to integrate US valence and validity information for negatively compared to positively paired CSs in the invalid condition (i.e., forming a CS-positive representation instead of a CS-negative-invalid representation). To control for possible valence asymmetries in the valid condition (which might go in the opposite direction), we consider the interaction effect between US valence and US validity instead of merely the main effect of US valence in the invalid condition. Note that the preregistrations failed to specify that this hypothesis refers to valence memory.

In line with this prediction, the valence memory measure in Experiment 2 showed a strong US valence × US validity interaction effect in the anterior condition, $F(1, 79) = 13.36$, $p < .001$, $\eta_p^2 = .145$, but not in the posterior condition, $F(1, 79) = 0.04$, $p = .850$, $\eta_p^2 = .000$. Indeed, the moderating effect of timing of validity information was also significant, $F(1, 158) = 8.11$, $p = .005$, $\eta_p^2 = .049$.

In Experiment 3, we observed similar effects although they were somewhat weaker. The valence memory measure (collapsed across sessions) showed a strong (US valence × validity information interaction) effect in the anterior condition, $F(1, 79) = 4.56$, $p = .036$, $\eta_p^2 = .145$, but not in the posterior condition, $F(1, 79) = 0.47$, $p = .497$, $\eta p^2 = .006$. The moderating effect of timing of validity information failed to reach significance, $F(1, 158) = 1.25$, $p = .266$, $\eta_p^2 = .008$.

Dependencies between MPT Parameters

(12) "Over time the u-parameter will decrease less strong[ly] in the postponed-instructions condition than in the advanced-instructions condition."

This hypothesis was derived from the notion of lost validity tags. As separate encoding should be facilitated when validity information is provided after the US valence is encoded, it can be expected that lost tags – and thereby an increase in evaluations that reflect merely US valence - will serve as a buffer against decreases of the *u*-parameter. However, a decelerated decrease alone does not indicate the source of this effect. If the *u*-parameter is fed by the forgetting of validity information, then it should show a correlation with the decrease over time observed in the *c*-parameter. The main article therefore focuses on correlational analyses that assess the dependency between the *c*- and *u*-parameters more directly.

Nevertheless, the assessment of this hypothesis is reported here. In Experiment 2, the *u*-parameter overall did not decrease significantly over the course of 24 hours, $F(1, 150) = 0.56$, $p = .454$, $\eta_p^2 = .004$. However, as expected, the change in parameter size depended on the timing of validity information, $F(1, 150) = 2.97$, $p = .043$, one-tailed, $\eta_p^2 = .019$. Whereas the *u*-parameter showed a decrease over time in the anterior validity condition, $F(1, 71) = 3.00$, $p = .088$, $\eta_p^2 = .041$, it was stable in the posterior condition, $F(1, 79) = 0.49$, $p = .488$, $\eta_p^2 = .006$ (descriptively it showed a slight increase from T1 to T2).

In Experiment 3, neither the decrease in the *u*-parameter over the course of 24 hours is significant, $F(1, 145) = 2.98$, $p = .086$, $\eta_p^2 = .020$, nor its dependency on the timing of validity information, $F(1, 145) = 1.40$, $p = .238$, $\eta_p^2 = .010$.

(13) "At Time 2, the u-parameter will be larger in the postponed-instructions condition than in the advanced-instructions condition indicating an influence of lost tags."

This hypothesis was based on the assumption that lost tags will artificially inflate the *u*-parameter especially in the posterior condition. On hindsight, we consider this hypothesis too strong. Its disconfirmation does not invalidate our demonstration that decreases in the *c*-parameter can artificially inflate the *u*-parameter when learning has taken place under conditions that foster separate encoding.

The hypothesis is not discussed in the main article and was indeed disconfirmed by the data (Exp. 2: $t(156) = -0.07$, $p = .947$; Exp. 3: $t(155) = 0.00$, $p = .996$). The main article instead presents correlational analyses performed on the changes of *c* and *u* over time.

References

Barnes, J. M., & Underwood, B. J. (1959). “Fate” of first-list associations in transfer theory. Journal of Experimental Psychology, 58(2), 97–105. https://doi.org/10.1037/h0047507

Ebbinghaus, H. (1885/1913). Memory: A contribution to experimental psychology. New York, NY, US: Teachers College Press.

Förderer, S., & Unkelbach, C. (2013). On the stability of evaluative conditioning effects. *Social Psychology, 44*(6), 380–389. https://doi.org/10.1027/1864-9335/a000150

Gast, A., De Houwer, J., & De Schryver, M. (2012). Evaluative conditioning can be modulated by memory of the CS–US pairings at the time of testing. *Learning and Motivation, 43*(3), 116–126. https://doi.org/10.1016/j.lmot.2012.06.001

Sweldens, S., Van Osselaer, S. M. J., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research, 37*(3), 473–489. https://doi.org/10.1086/653656

Supplement S2

Instructions to Participants

Instructions Used in Experiments 1 and 2 (Control Paradigm)

Every day new and unfamiliar brands are introduced to the market. Previous research on advertising has shown that we quickly form impressions of unknown brands and that this impression strongly depends on the context in which we encounter a brand. Specifically, this means that we tend to like brands that we encounter in a positive context and that we tend to dislike brands that we encounter in a negative context.

In advertising, brands are often presented in a positive or negative context. Sometimes these contexts can be informative about a brand. High quality marketing with very positive images can indicate a high-quality product, while marketing with low-quality and negative images can accordingly point to inferior products. At the same time, such contexts can readily give a false impression. For instance, products can be of low quality despite a favorable presentation, or high-quality products can be presented in an unfavorable light by other brands.

Anterior Validity Information

In this research program, we examine whether people are able to use contextual information when appropriate and to reverse the influence of the context when it conflicts with the quality or characteristics of a brand.

In this study you will be presented with logos of drinking water brands along with positive and negative images. We will also let you know whether the positive or negative images provide correct or incorrect information about the water brands. This is done in two phases.

In one phase the images will be correct. Accordingly, you should apply their valence to form a correct impression of the water brands.

Thus: you should start liking brands whose logos are presented together with positive images; you should start disliking brands whose logos are presented together with negative images.

In another phase the pictures will be incorrect. Accordingly, you should reverse their valence in order to get a correct impression of the water brands.

Thus: you should start disliking brands whose logos are presented together with positive images; you should start liking brands whose logos are presented together with negative images.

In a later phase of this study, we will ask you how you feel about each water brand.

Valid Condition. Now you will [again] take part in a simple perception task. In this task, various logos of drinking water brands will be briefly presented together with positive or negative on the screen. Each logo will be presented several times for a few seconds together with a positive or negative context picture.

In this phase, the valent images provide correct information about the water brands. Thus, you should APPLY the valence of the pictures to form a correct impression of the water brands.

If a logo was presented with POSITIVE images, you should start LIKING the brand.

If a logo was presented with NEGATIVE images, you should start DISLIKING the brand.

Invalid Condition. Now you will [again] take part in a simple perception task. In this task, various logos of drinking water brands will be briefly presented together with positive or negative on the screen. Each logo will be presented several times for a few seconds together with a positive or negative context picture.

In this phase, the valent images provide incorrect information about the water brands. Thus, you should REVERSE the valence of the pictures to form a correct impression of the water brands.

If a logo was presented with POSITIVE images, you should start DISLIKING the brand.

If a logo was presented with NEGATIVE images, you should start LIKING the brand.

Posterior Validity Information

In this research program, we examine whether people are able to use contextual information when appropriate and to reverse the influence of the context when it conflicts with the quality or characteristics of a brand.

In this study you will be presented with logos of drinking water brands along with positive and negative images. We will also let you know whether the positive or negative images provide correct or incorrect information about the water brands. This is done in two phases.

In one phase the images will be correct. Accordingly, you should apply their valence to form a correct impression of the water brands.

Thus: you should start liking brands whose logos are presented together with positive images; you should start disliking brands whose logos are presented together with negative images.

In another phase the pictures will be incorrect. Accordingly, you should reverse their valence in order to get a correct impression of the water brands.

Thus: you should start disliking brands whose logos are presented together with positive images; you should start liking brands whose logos are presented together with negative images.

However, we will tell you only AFTER THE TWO PERCEPTION PHASES, which phase requires that you APPLY the valent images and which phase requires that you REVERSE the valent images in order to form a correct impression of the water brands.

In a later phase of this study, we will ask you how you feel about each water brand.

Instructions for the two conditioning phases. Now you will [again] take part in a simple perception task. In this task, various logos of drinking water brands will be briefly presented together with positive or negative on the screen. Each logo will be presented several times for a few seconds together with a positive or negative context picture.

ATTENTION: We will tell you only after the completion of the two perception phases whether you have to APPLY or REVERSE the valent pictures in this phase to form a correct impression of the water brands.

Application instructions provided after conditioning. In the first [second] perception phase, you were presented with various logos of drinking water brands together with positive or negative images.

In this first [second] phase, the valent images provided correct information about the water brands. Thus, you should APPLY the valence of the pictures to form a correct impression of the water brand.

If a logo was presented with POSITIVE images, you should start LIKING the brand.

If a logo was presented with NEGATIVE images, you should start DISLIKING the brand.

On the next screen we will again show you all the drinking water brands that were presented in the first [second] phase. This screen will be shown to you for 2 minutes. As soon as the 2 minutes have elapsed, the study continues automatically.

Reversal instructions provided after conditioning. In the first [second] perception phase, you were presented with various logos of drinking water brands together with positive or negative images.

In this first [second] phase, the valent images provided incorrect information about the water brands. Thus, you should REVERSE the valence of the pictures to form a correct impression of the water brand.

If a logo was presented with POSITIVE images, you should start DISLIKING the brand.

If a logo was presented with NEGATIVE images, you should start LIKING the brand.

On the next screen we will again show you all the drinking water brands that were presented in the first [second] phase. This screen will be shown to you for 2 minutes. As soon as the 2 minutes have elapsed, the study continues automatically.

Instructions Used in Experiment 3 (Relational Paradigm)

Every day new companies are founded. Accordingly, we encounter new brands almost every day. An increasingly important topic is the sustainability that characterizes a brand. Some companies set themselves high goals so as not to pollute the environment or to compensate for environmental damage through sustainability measures. Others exploit the environment for their own profit. Hence, there are sustainably operating companies on the one hand and exhaustingly operating companies on the other.

In this study we will introduce you to various companies:

SUSTAINABLY operating companies:

(1) Companies that PROMOTE ENVIRONMENTAL PROTECTION (e.g., through programs compensating for environmental damage)

(2) Companies that PREVENT ENVIRONMENTAL DESTRUCTION (e.g., by avoiding the emission of pollutants)

UNSUSTAINIBLY operating companies:

(3) Companies that PROMOTE ENVIRONMENTAL DESTRUCTION (e.g., through the emission of pollutants)

(4) Companies that PREVENT ENVIRONMENTAL PROTECTION (e.g., by rejecting programs compensating for environmental damage)

Anterior Validity Information

In this study, we will show you different brand logos together with positive images of intact and attractive landscapes and negative images of destroyed and unattractive landscapes. The attractive landscapes symbolize environmental protection. The unattractive landscapes symbolize environmental destruction.

We will also inform you whether the company behind the brand promotes or prevents the condition shown in the pictures. This is done in two phases:

In one phase, the companies behind the brands PROMOTE the respective condition.

In another phase, the companies behind the brands PREVENT the respective condition.

As a reminder, we distinguish between:

SUSTAINABLY operating companies:

(1) Companies that PROMOTE ENVIRONMENTAL PROTECTION (e.g., through programs compensating for environmental damage)

(2) Companies that PREVENT ENVIRONMENTAL DESTRUCTION (e.g., by avoiding the emission of pollutants)

UNSUSTAINABLY operating companies:

(3) Companies that PROMOTE ENVIRONMENTAL DESTRUCTION (e.g., through the emission of pollutants)

(4) Companies that PREVENT ENVIRONMENTAL PROTECTION (e.g., by rejecting programs compensating for environmental damage)

Valid condition. Now you will perform the first [second] part of this task. In this task, various logos of drinking water brands will be briefly presented together with images of attractive or unattractive landscapes. Each logo will be presented several times for a few seconds together with an image of an attractive or unattractive landscape.

In this phase, we present you only with brands that PROMOTE the depicted condition. Thus, behind the brand logos are companies that promote environmental protection and companies that promote environmental destruction. Specifically, this means:

When a brand is presented with ATTRACTIVE LANDSCAPES, this means that the company behind it PROMOTES ENVIRONMENTAL PROTECTION.

When a brand is presented with UNATTRACTIVE LANDSCAPES, this means that the company behind it PROMOTES ENVIRONMENTAL DESTRUCTION.

Invalid condition. Now you will perform the first [second] part of this task. In this task, various logos of drinking water brands will be briefly presented together with images of attractive or unattractive landscapes. Each logo will be presented several times for a few seconds together with an image of an attractive or unattractive landscape.

In this phase, we present you only with brands that PREVENT the depicted condition. Thus, behind the brand logos are companies that prevent environmental protection and companies that prevent environmental destruction. Specifically, this means:

When a brand is presented with ATTRACTIVE LANDSCAPES, this means that the company behind it PREVENTS ENVIRONMENTAL PROTECTION.

When a brand is presented with UNATTRACTIVE LANDSCAPES, this means that the company behind it PREVENTS ENVIRONMENTAL DESTRUCTION.

Posterior Validity Information

In this study, we will show you different brand logos together with positive images of intact and attractive landscapes and negative images of destroyed and unattractive landscapes. The attractive landscapes symbolize environmental protection. The unattractive landscapes symbolize environmental destruction.

We will also inform you whether the company behind the brand promotes or prevents the condition shown in the pictures. This is done in two phases:

In one phase, the companies behind the brands PROMOTE the respective condition.

In another phase, the companies behind the brands PREVENT the respective condition.

As a reminder, we distinguish between:

SUSTAINABLY operating companies:

(1) Companies that PROMOTE ENVIRONMENTAL PROTECTION (e.g., through programs compensating for environmental damage)

(2) Companies that PREVENT ENVIRONMENTAL DESTRUCTION (e.g., by avoiding the emission of pollutants)

UNSUSTAINABLY operating companies:

(3) Companies that PROMOTE ENVIRONMENTAL DESTRUCTION (e.g., through the emission of pollutants)

(4) Companies that PREVENT ENVIRONMENTAL PROTECTION (e.g., by rejecting programs compensating for environmental damage)

However, we will only tell you AFTER THE COMPLETION OF BOTH PHASES, in which phase the companies PROMOTE the depicted condition and in which phase the companies PREVENT the depicted condition.

Instructions for the two conditioning phases. Now you will [again] perform a simple perception task. In this task, various brand logos will be briefly presented together with images of attractive or unattractive landscapes. Each logo will be presented several times for a few seconds together with an image of an attractive or unattractive landscape.

ATTENTION: We will only tell you AFTER THE COMPLETION OF BOTH PHASES, in which phase the companies PROMOTE the depicted condition and in which phase the companies PREVENT the depicted condition.

Application instructions provided after conditioning. In the first [second] phase, various brand logos of companies were presented together with attractive or unattractive landscapes on the screen.

The companies presented in this first [second] phase PROMOTE the depicted condition.

When a brand was presented with ATTRACTIVE LANDSCAPES, this means that the company behind it PROMOTES ENVIRONMENTAL PROTECTION.

When a brand was presented with UNATTRACTIVE LANDSCAPES, this means that the company behind it PROMOTES ENVIRONMENTAL DESTRUCTION.

On the next screen we will again show you all the brand logos that were presented in the first [second] phase. This screen will be shown to you for 2 minutes. As soon as the 2 minutes have elapsed, the study continues automatically.

Reversal instructions provided after conditioning. In the first [second] phase, various brand logos of companies were presented together with attractive or unattractive landscapes on the screen.

The companies presented in this first [second] phase PREVENT the depicted condition.

When a brand was presented with ATTRACTIVE LANDSCAPES, this means that the company behind it PREVENTS ENVIRONMENTAL PROTECTION.

When a brand was presented with UNATTRACTIVE LANDSCAPES, this means that the company behind it PREVENTS ENVIRONMENTAL DESTRUCTION.

On the next screen we will again show you all the brand logos that were presented in the first [second] phase. This screen will be shown to you for 2 minutes. As soon as the 2 minutes have elapsed, the study continues automatically.

Supplement S3

Performance on the valence and validity memory measures

	Experiment 1				Experiment 2								Experiment 3							
					anterior information				posterior information				anterior information				posterior information			
	valid pairings		invalid pairings		valid pairings		invalid pairings		valid pairings		invalid pairings		promotion CSs		prevention CSs		promotion CSs		prevention CSs	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
valence and validity correct	10.52 (3.94)	8.81 (4.14)	6.95 (4.65)	5.44 (4.03)	11.00 (4.05)	9.51 (3.67)	7.46 (4.10)	6.49 (3.45)	8.98 (3.83)	7.96 (3.31)	8.93 (3.70)	6.84 (3.61)	10.55 (3.54)	8.86 (3.13)	7.64 (3.97)	6.38 (3.62)	7.84 (3.55)	7.35 (3.61)	7.24 (3.64)	6.26 (3.22)
valence and validity incorrect	0.95 (1.35)	1.38 (1.61)	2.08 (1.84)	2.79 (2.07)	1.88 (2.23)	2.59 (2.15)	2.64 (2.01)	3.58 (2.35)	1.59 (1.83)	2.34 (1.76)	1.44 (1.43)	1.96 (1.89)	2.55 (2.02)	3.81 (2.39)	4.55 (3.26)	5.73 (3.13)	2.35 (2.28)	2.89 (2.14)	2.28 (2.27)	3.03 (2.34)
valence incorrect	0.89 (1.12)	0.82 (1.18)	2.63 (2.47)	1.94 (1.89)	1.48 (1.81)	1.91 (1.82)	3.56 (3.00)	3.09 (2.32)	2.36 (2.20)	2.39 (2.03)	1.93 (1.59)	2.46 (1.84)	1.45 (1.81)	1.60 (2.00)	1.66 (1.99)	1.53 (1.74)	2.69 (2.25)	3.06 (2.34	2.59 (1.93)	2.69 (1.79)
validity incorrect	1.24 (1.78)	1.37 (1.98)	1.67 (2.27)	1.61 (2.31)	1.65 (1.81)	1.99 (1.83)	2.34 (2.35)	2.85 (2.46)	3.08 (2.10)	3.31 (2.08)	3.71 (2.52)	4.74 (2.56)	1.45 (2.09)	1.73 (2.36)	2.15 (2.90)	2.38 (2.99)	3.13 (2.23)	2.71 (2.30)	3.90 (2.79)	4.03 (2.79)
‘don’t know’	2.41 (2.84)	3.61 (4.04)	2.67 (2.70)	4.30 (3.98)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Note. Cell entries denote mean frequencies, values in brackets denote standard deviations.

Supplement S4

Frequency Data Underlying MPT Analyses

Aggregated Observed Frequencies in the Pleasant (+) and Unpleasant (–) Response Categories as a Function of US Valence, Instruction Condition, and Timing of Instructions.

		Valid Condition		Invalid Condition		
		+	–	+	–	**Σ observations**
Experiment 1 [Anterior Validity Information]						
Session 1	CS+	548	84	176	456	
	CS–	85	547	418	214	
Session 2	CS+	483	149	235	397	
	CS–	135	497	380	252	5056
Experiment 2: Anterior Validity Information Condition						
Session 1	CS+	521	119	203	437	
	CS–	111	529	409	231	
Session 2	CS+	466	174	250	390	
	CS–	186	440	349	291	
Experiment 2: Posterior Validity Information Condition						
Session 1	CS+	432	208	264	376	
	CS–	165	475	366	274	
Session 2	CS+	406	234	349	291	
	CS–	200	440	253	387	10240
Experiment 3: Anterior Validity Information Condition						
Session 1	CS+	513	127	155	485	
	CS–	90	550	472	168	
Session 2	CS+	496	144	174	466	
	CS–	119	521	434	206	
Experiment 3: Posterior Validity Information Condition						
Session 1	CS+	429	211	284	356	
	CS–	196	444	328	312	
Session 2	CS+	427	213	306	334	
	CS–	218	422	310	330	10240

Note. “CS+” denote CSs paired with positive USs; “CS-“ denote CSs paired with negative USs.