Mathematical Characterization of Changes in Fear During Exposure Therapy

Ana Portêlo, Youssef Shiban, and Tiago V. Maia

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL TEXT

ANALYSES USING CLASSICAL STATISTICS

Methods

We first analyzed the data using classical statistics—specifically, mixed ANOVAs. When the sphericity assumption was violated, we used Greenhouse-Geisser correction (1).

Fear Decrease

We defined fear decrease within an exposure as the difference between the fear ratings at the beginning and end of the exposure. We tested if there was evidence for fear decrease within exposures in each group by using a multivariate one-sample Hotelling's T^2 test comparing fear decrease for the four training exposures with the vector [0 0 0 0] (one entry per exposure), separately for each group. We then used a mixed ANOVA to analyze the effects of the between-subjects factor Group (single context, multiple contexts) and the within-subjects factor Exposure (0, 1, 2, 3). (We number exposures starting at 0 because that is more convenient mathematically for the model-based analyses.)

Return of Fear

We defined return of fear between consecutive exposures as the difference between fear ratings at the beginning of a given exposure and at the end of the previous exposure. We started by testing if there was evidence for return of fear in each group by using a multivariate one-sample Hotelling's T^2 test comparing return of fear for the three intervals between consecutive exposures with the vector [0 0 0] (one entry per interval), separately for each group. We then used a mixed ANOVA to analyze the effects of the between-subjects factor Group (single context, multiple contexts) and the within-subjects factor Exposure Interval, defined between consecutive exposures (<0, 1>, <1, 2>, <2, 3>).

Results

Fear Decrease

Fear decrease within exposures

Both groups exhibited fear decrease within exposures (Figure 2A; one-sample Hotelling's T^2 test of fear decrease for the four training exposures against [0 0 0 0] for the single-context group: $F = 16.04$, $df = 4, 11$, $p < .001$; mean fear decrease for exposure

0 = 17.67, 95% CI [14.11, 21.22]; mean fear decrease for exposure 1 = 18.93, 95% CI [16.22, 21.64]; mean fear decrease for exposure 2 = 11.33, 95% CI [8.53, 14.14]; mean fear decrease for exposure 3 = 7.47, 95% CI [4.54, 10.39]; one-sample Hotelling's T^2 test of fear decrease for the four training exposures against [0 0 0 0] for the multiple-contexts group: *F* = 18.88, *df* = 4, 11, *p* < .001; mean fear decrease for exposure 0 = 39.00, 95% CI [33.65, 44.35]; mean fear decrease for exposure 1 = 27.40, 95% CI [22.98, 31.82]; mean fear decrease for exposure 2 = 23.33, 95% CI [19.48, 27.19]; mean fear decrease for exposure 3 = 15.13, 95% CI [11.21, 19.06]). Fear decrease within exposures was greater in the multiple-contexts group than in the single-context group (Figure 2A; main effect of Group in the mixed ANOVA: *F* = 9.29, *df* = 1, 28, *p* = .005, η^2 = .249).

Change in fear decrease across exposures

Fear decrease within exposures changed across exposures (Figure 2A; main effect of Exposure in the mixed ANOVA: *F* = 9.55, *df* = 2.17, 60.67, *p* <. 001, η^2 = .254), with no evidence that this change differed between the groups (interaction of Group × Exposure in the mixed ANOVA: *F* = 1.73, *df* = 2.17, 60.67, *p* = .183, η^2 = .058). The change in fear decrease across exposures exhibited a significant linear trend for the two groups combined (linear contrast: *F* = 20.383, *df* = 1, 28, *p* = .001, η^2 = .421) and for each group separately (linear contrast for the single-context group: *F* = 9.75, *df* = 1, 14, *p* = .007, η^2 = .411; linear contrast for the multiple-contexts group: *F* = 11.77, *df* = 1, 14, *p* = .004, η^2 = .457). Although fear decrease for the single-context group did not show a strict monotonic decrease across exposures (Figure 2A right), we found no significant evidence of nonmonotonicity using a quadratic contrast (*F* = 0.72, *df* = 1, 14, *p* = .410, η^2 = .049). Failure to reject the null hypothesis cannot, of course, be construed as evidence for the null hypothesis, but, overall, these results suggest that the amount of fear decrease diminishes across exposures in both groups.

Return of Fear

Return of fear between consecutive exposures

The multiple-contexts group exhibited return of fear between consecutive exposures (Figure 2B; one-sample Hotelling's T^2 test of return of fear for the three exposure intervals against [0 0 0]: *F* = 5.64, *df* = 3, 12, *p* = .007; mean return of fear for the exposure interval <0, 1> = 18.00, 95% CI [14.35, 21.65]; mean return of fear for the exposure interval <1, 2> = 18.73, 95% CI [15.43, 22.03]; mean return of fear for the exposure interval <2, 3> = 11.13, 95% CI [9.40, 12.86]), but there was no evidence for return of fear in the single-context group (Figure 2B; one-sample Hotelling's T^2 test of return of fear for the three exposure intervals against [0 0 0]: *F* = 0.59, *df* = 3, 12, *p* = .602; mean return of fear for the exposure interval <0, 1> = –2.07, 95% CI [–4.15, 0.02]; mean return of fear for the exposure interval <1, 2> = –2.07, 95% CI [–3.81, –0.32]; mean return of fear for the exposure interval <2, 3> = 0.87, 95% CI [–0.90, 2.63]). Return of fear was significantly greater in the multiple-contexts than in the single-context group (Figure 2B; main effect of Group in the mixed ANOVA: *F* = 14.11, *df* = 1, 28, *p* = .001, η^2 = .335).

Change in return of fear across exposure intervals

There was no evidence that return of fear changed across exposure intervals (main effect of Exposure Interval in the mixed ANOVA: *F* = 0.39, *df* = 2, 56, *p* = .681, η^2 = .014) nor that any such potential changes varied by group (interaction of Group × Exposure Interval in the mixed ANOVA: *F* = 2.13, *df* = 2, 56, *p* = .129, η^2 = .071). Although the plot of return of fear for the multiple-contexts group seemed to suggest that return of fear in the third exposure interval might be smaller than those in the first and second exposure intervals (Figure 2B right), that difference was not statistically significant ([1/2 1/2 –1] contrast: *F* = 3.00, *df* = 1, 14, *p* = .105, η^2 = .176).

Discussion

Fear Decrease

Both groups exhibited fear decrease within exposures, with fear decrease being stronger in the multiple-contexts group. Fear decrease diminished across exposures, with no evidence that this reduction differed between the groups. The reduction in fear decrease across exposures may be due to some effect of repetition; alternatively, it may simply reflect the higher initial fear ratings in earlier relative to later exposures (Figure 2A and Figure S6). Higher initial fear ratings allow more fear decrease before ratings run against a floor effect (ratings of 0). Moreover, higher initial fear ratings *imply* more fear decrease, even independently of floor effects, if the fear-decrease rate is proportional to the level of fear (as would occur if fear decays exponentially within exposures). Simple statistical analyses are insufficient to adjudicate between these alternatives; we address them in full, however, with the model-based analyses.

The group difference in fear decrease may reflect differential effects of the treatment protocols or, as for the reduction in fear decrease across exposures, it may simply reflect differences in initial fear ratings: these ratings are slightly higher in the multiple-contexts relative to the single-context group (for exposures 2 and 3) due to return of fear in the multiple-contexts but not in the single-context group (Figure 2A and Figure S6). A third and perhaps more likely possibility is that this difference reflects an unfortunate aspect of this dataset: in the very first exposure, when the protocol is exactly equal for the two groups, the multiple-contexts group already shows markedly and statistically significantly greater fear decrease than the single-context group does (Figure 2A; independent-samples *t*-test comparing fear decrease between the two groups in the first exposure: *t* = 3.08, *df* = 28, *p* = .005, mean difference of multiple-contexts group minus single-context group = 21.33, 95% CI [7.13, 35.54]). This finding cannot be attributed to treatment effects, so it likely reflects a failure of randomization to balance the groups in this aspect due to the relatively small samples in each group (*n* = 15 in each group). For this reason, we will not address further the group difference in fear decrease.

Return of Fear

The groups differed significantly in return of fear, and there was evidence for return of fear in the multiple-contexts but not in the single-context group. This finding was expected: return of fear typically occurs through reinstatement, reacquisition, spontaneous recovery, or renewal (2,3). The exposure-therapy protocol used did not provide opportunities for reinstatement or reacquisition, and spontaneous recovery likely was negligible given the close temporal proximity between exposures (2 min). Return of fear therefore occurred in the multiple-contexts group, which elicited renewal by changing the exposure context, but not in the single-context group, given that renewal is minimal when the context does not change (2,3).

We found no statistically significant evidence that return of fear changed across exposure intervals, even in the multiple-contexts group. We suspect that, for the multiple-contexts group, this null result was due to the small number of exposure intervals tested (just 3): prior studies have found decreased renewal following extinction in different contexts (4,5) but only after a sufficient number of such extinctions (6). Indeed, considering the additional interval between the last training exposure and the test exposure suggests that return of fear might decrease across exposure intervals in the multiple-contexts group (Figure S8).

MODEL-BASED ANALYSES

Mathematical Formulation of the Models

Structure of the Full Models

We aimed to simultaneously capture each subject's full set of fear ratings. Let $F(x, t)$ represent the observed fear rating at time t for exposure x, where t represents time within the exposure (0, 1, 2, 3, or 4, because there were 5 ratings per exposure) and x represents the exposure number (0, 1, 2, or 3 for the four training exposures; 4 for the test exposure). We start numbering at 0 for mathematical convenience. We omit the subject number for simplicity, but all functions refer to an individual subject. Our goal was to find the best model for $F(x, t)$. (More precisely, we fit the model only to the training exposures; we used the test exposure to test the model's predictions.)

We modeled the processes that contribute to $F(x, t)$ using three functions: D, S, and R

(Figure 3). Function $D_{[\hat{F}(x,\,0),\,S(x)]}(t)$ modeled fear decrease within exposure x. Generally, D was a decreasing (or possibly constant) function of time t within the exposure. This function was characterized by two parameters specific to exposure x: the estimated fear at the beginning of the exposure [$\hat{F}(x, 0)$] and a parameter determining the fear-decrease steepness for the exposure [$S(x)$]. These parameters were not directly estimated by the model; instead, they were determined by the interaction of the various model components, as described below.

The estimated initial fear rating for each exposure, $\hat{F}(x, 0)$, was not determined by D but rather was a parameter of D. We need to distinguish between the estimated initial fear rating for the first exposure [$\hat{F}(0, 0)$] and that for all other exposures [$\hat{F}(x, 0)$, with $x > 0$] because the latter but not the former are affected by preceding fear decrease within exposures and return of fear between exposures. For $\hat{F}(0, 0)$, we considered using the observed initial fear rating [$\hat{F}(0, 0) = F(0, 0)$] or having $\hat{F}(0, 0)$ as a free parameter. Preliminary model-selection analyses (not shown) demonstrated that having $\hat{F}(0, 0)$ as a free parameter did not improve the fit sufficiently to justify the additional model complexity; we therefore made $\hat{F}(0, 0) = F(0, 0)$. For all other exposures ($x > 0$), we defined the estimated initial fear, $\hat{F}(x, 0)$, as the sum of the estimated fear at the end of the prior exposure, $\hat{F}(x-1, 4)$, and the estimated return of fear between the exposures, $R(<x-1, x>)$:

$$\hat{F}(x, 0) = \hat{F}(x-1, 4) + R(\langle x-1, x\rangle). \tag{SE1}$$

We will define functions $R(<x-1, x>)$ and $S(x)$ in more detail below. For now, suffice it so say that $R(<x-1, x>)$ gives the return of fear from exposures $x-1$ to x, and $S(x)$ determines the steepness of D for exposure x.

Together, functions D, S, and R determine the complete model:

$$\hat{F}(x, t) = \begin{cases} F(0, 0), & \text{if } x = 0,\ t = 0; \\ \hat{F}(x-1, 4) + R(\langle x-1, x\rangle), & \text{if } x > 0,\ t = 0; \\ D_{[\hat{F}(x,\,0),\,S(x)]}(t), & \text{if } t > 0. \end{cases} \tag{SE2}$$

At this level, the model does not have any free parameters; those are embedded in functions S and R, as described below.

Model Components

Fear Decrease Within Exposures

We tested the following formulations for $D_{[\hat{F}(x,\,0),\,S(x)]}(t)$:

$$D^{\text{Const}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x, 0); \tag{SE3}$$

$$D^{\text{Lin}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x, 0) - \mathrm{S}(x)\,t; \tag{SE4}$$

$$D^{\mathrm{Exp}}_{[\hat{F}(x,\,0),\,\mathrm{S}(x)]}(t) = \hat{F}(x,\,0)\ \mathrm{e}^{-S(x)\,t}; \tag{SE5}$$

$$D^{\mathrm{Ln}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x,\,0) - S(x)\ln(1+t); \tag{SE6}$$

$$D^{\mathrm{Pow}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x,\,0)\ (1+t)^{-S(x)}; \tag{SE7}$$

$$D^{\mathrm{LinSqrt}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x,\,0) - S(x)\sqrt{t}; \tag{SE8}$$

$$D^{\mathrm{ExpSqrt}}_{[\hat{F}(x,\,0),\,S(x)]}(t) = \hat{F}(x,\,0)\ \mathrm{e}^{-S(x)\sqrt{t}}. \tag{SE9}$$

The superscript identifies the shape of the function in terms of t: constant (Const), linear (Lin), exponential (Exp), logarithmic (Ln), power (Pow), linear on the square root (LinSqrt), and exponential on the square root (ExpSqrt). D^{Const} assumes no fear decrease within exposures: fear is constant and equal to its initial value for the exposure [$\hat{F}(x, 0)$] (Equation SE3); this function therefore ignores $S(x)$, which has no meaning if there is no fear decrease. Alternatively, D^{Const} can be seen as a special case of the other functions with $S(x) = 0$ for all x. All other functions (Equations SE4–SE9) define monotonically decreasing functions of t if $S(x)$ is positive. None of the functions have any free parameters; their parameters (in subscripted square brackets) are determined by an interaction of the various model components.

As mentioned in the main text, we had some *a priori* expectation that the exponential function might be best at describing fear decrease within exposures. However, all other functions that we tested also had some theoretical justification. Specifically, we tested the power function because of the suggested power law of practice (7,8); we tested the linear function because, in some cases, habituation may be best fit linearly (9); we tested the constant function as a base case, to be ruled out, of no fear decrease; we tested the three other functions (linear on the logarithm, linear on the square root, and exponential on the square root) because they have been used to describe various learning and forgetting processes (10,11).

Some of these functions (D^{Exp}, D^{Pow}, and D^{ExpSqrt}) asymptote at 0, which seems more plausible because fear ratings cannot go below 0. Still, we tested functions that can become negative (D^{Lin}, D^{Ln}, and D^{LinSqrt}) because they could better capture the data in the relevant range. There is even some theoretical justification for functions that represent fear (even if not fear ratings) to become negative because overtraining in both habituation (9) and extinction (12) produces effects that continue to strengthen beyond response cessation.

Change in Fear-Decrease Steepness Across Exposures

Function $S(x)$ gives a coefficient that determines the steepness of D for exposure x [see the role of $S(x)$ in Equations SE4–SE9]. We tested the following formulations for $S(x)$:

$$S^{\mathrm{Const}}(x) = \lambda; \tag{SE10}$$

$$S^{\mathrm{Lin}}(x) = \lambda_0 + \lambda_1 x. \qquad \text{(SE11)}$$

The lambdas were free parameters.

Return of Fear

Function $R(<x-1, x>)$ gives the estimated return of fear between exposures $x-1$ and x (with $x > 0$). We tested the following formulations for $R(<x-1, x>)$:

$$R^{\mathrm{No}}(\langle x-1, x\rangle) = 0; \qquad \text{(SE12)}$$

$$R^{\mathrm{Const}}(\langle x-1, x\rangle) = \rho; \qquad \text{(SE13)}$$

$$R^{\mathrm{Prop}}(\langle x-1, x\rangle) = \alpha\, [\hat{F}(x-1, 0) - \hat{F}(x-1, 4)]. \qquad \text{(SE14)}$$

Function R^{No} had no parameters; R^{Const} had a single parameter (ρ) that corresponded to the amount of constant return of fear; R^{Prop} also had a single parameter (α, expected to be between 0 and 1) that defined the proportion of fear that had decreased in the previous exposure that returned at the beginning of the present exposure. We did not test more complex formulations for R because of the small number of exposure intervals.

Free Parameters

In summary, the models had between zero and three free parameters: D had no free parameters; S had one (λ) or two (λ_0 and λ_1) free parameters [or, in the special case of no fear decrease (Equation SE3), S was not part of the model, so there were no parameters for S]; R had zero or one (ρ or α) free parameter.

Chosen Models

As shown in the main text, the chosen model for both groups had exponentially decaying fear within exposures (D^{Exp}; Equation SE5) with a constant decay rate (S^{Const}; Equation SE10). The groups differed in their return-of-fear model: the single-context group had constant return of fear (R^{Const}; Equation SE13) and the multiple-contexts group had proportional return of fear (R^{Prop}; Equation SE14). Putting together the corresponding equations, we obtain the following:

$$\hat{F}(x, t) = \begin{cases} F(0, 0), & \text{if } x = 0,\ t = 0; \\ \hat{F}(x-1, 4) + R(<x-1, x>), & \text{if } x > 0,\ t = 0; \\ \hat{F}(x, 0)\, \mathrm{e}^{-\lambda t}, & \text{if } t > 0. \end{cases} \qquad \text{(SE15)}$$

The top line in the braces sets the estimated initial fear for the first exposure to its observed value; the middle line defines the jumps that occur between exposures due to return of fear; the bottom line describes the exponentially decreasing fear within exposures. Function R differed between the groups:

$$R(<x-1,x>)=\begin{cases}\rho, & \text{if Group = SC;}\\ \alpha\,[\hat{F}(x-1,0)-\hat{F}(x-1,4)], & \text{if Group = MC;}\end{cases} \tag{SE16}$$

where SC and MC represent the single-context and multiple-context groups, respectively. The exponential decay captured in the bottom line of Equation SE15 is the solution to the following first-order homogeneous linear differential equation with constant coefficients:

$$\frac{\mathrm{d}\,\hat{F}(x,t)}{\mathrm{d}t}=-\lambda\,\hat{F}(x,t), \tag{SE17}$$

with $\hat{F}(x, 0)$ as the initial value. For the first exposure ($x = 0$), this initial value is given by the observed value [$F(0, 0)$]; for all subsequent exposures, the initial value is given by a jump defined by Equation SE16. This model is therefore a hybrid dynamical system: a system that combines continuous dynamics (within exposures) with discrete dynamics (between exposures) (13). For both groups, the chosen model had only two free parameters: λ and ρ for the single-context group, and λ and α for the multiple-contexts group.

Additional Information about Model Fitting and Selection

We fit each candidate model as a mixed-effects model to all fear ratings of all subjects from each group, using MATLAB's nlmefit function. We included random effects for all parameters to obtain subject-specific parameters. The model-fitting process resulted in a Bayesian Information Criterion (BIC) value for each subject-model pair; we used the set of BIC values for all subject-model pairs as inputs for Bayesian model selection (14,15).

As noted in the main text, we used the Variational Bayesian Analysis (VBA) toolbox's ability to group models into families. We used uniform priors over the families and, within each family, over all models in the family.

Accuracy of the Model Fits and Predictions

Quantitative analyses demonstrated the excellence of the model fits for the training exposures in both groups and of the model predictions for the test exposure in the multiple-contexts group. (As discussed in the main text, we did not assess model predictions in the test exposure for the single-context group because we did not expect the model for that group to generalize well to the test exposure.) The mean absolute value of the residuals, averaged across subjects, was remarkably low in all cases (training exposures for the single-context group: mean = 5.06, *SD* = 1.21, median = 5.23; training exposures for the multiple-contexts group: mean = 6.51, *SD* = 3.02, median = 5.94; test exposure for the multiple-contexts group: mean = 6.32, *SD* = 5.61, median = 5.44; Table

S2), indicating that the average prediction error as a percentage of the full rating-scale range (0–100) was always in the mid-single digits. The root mean square error, averaged across subjects, gave similar results (training exposures for the single-context group: mean = 6.11, *SD* = 1.29, median = 6.34; training exposures for the multiple-contexts group: mean = 7.63, *SD* = 3.27, median = 6.87; test exposure for the multiple-contexts group: mean = 7.03, *SD* = 6.12, median = 6.34; Table S2). Furthermore, the correlation between predicted and observed values, averaged across subjects, was remarkably high in all cases (training exposures for the single-context group: mean = .93, *SD* = 0.06, median = .94; training exposures for the multiple-contexts group: mean = .86, *SD* = 0.14, median = .91; test exposure for the multiple-contexts group: mean = .87, *SD* = 0.16, median = .92; Table S2), suggesting that the shapes of the fitted and observed curves were similar (as can also be seen by visual inspection; Figure 5). Finally, R^2, calculated by comparing the fits of our model against those of a null model consisting of a constant value (see below), was also remarkably high for the training exposures (single-context group: mean = 0.87, *SD* = 0.11, median = 0.88; multiple-contexts group: mean = 0.81, *SD* = 0.21, median = 0.83; Table S2) and for most subjects from the multiple-contexts group in the test exposure (median = 0.83). The mean R^2 for subjects from the multiple-contexts group in the test exposure was not as high (mean = 0.48, *SD* = 0.96), but this was largely due to two subjects whose behavior was better predicted by the null model: subjects #34 and #51 (Table S2). These subjects exhibited atypical behavior, with fear increases mid-way through the test exposure (Figure 5B), which explains why their fear ratings were not well predicted by the model. Excluding these two subjects, the mean R^2 for subjects from the multiple-contexts group in the test exposure was also remarkably high (mean = 0.78), with a much lower *SD* (*SD* = 0.28). Graphical analysis of the residuals further demonstrates the excellent quality of the fit: residuals show no systematic deviations from model predictions at any time point, and standardized residuals are strongly concentrated around 0, with a normal distribution, and with no autocorrelation (Figure S3).

Calculation of R^2

We calculated R^2 for each subject *i*, which we represent here by R_i^2, by comparing the fits of our model against those of a null model consisting of a constant value. Specifically, we used the following formula:

$$R_i^2 = 1 - \frac{\sum_{x=x_0}^{X}\sum_{t=0}^{T-1}\left(D_{i,x,t} - P_{i,x,t}\right)^2}{\sum_{t=0}^{T-1}\left(D_{i,x,t} - \bar{D}\right)^2}, \qquad \text{(SE18)}$$

where *x* represents an exposure; x_0 and *X* represent, respectively, the minimum and maximum indices of the exposures under consideration; *t* represents time within an exposure; *T* represents the number of time points per exposure; $D_{i,\,x,\,t}$ and $P_{i,\,x,\,t}$

represent, respectively, the observed and predicted fear ratings at time *t* in exposure *x* for subject i; and $\overline{D}$ represents the overall average of the observed fear ratings at all time points for the exposures under consideration for all subjects from the group under consideration:

$$\overline{D} = \frac{\sum_i \sum_{x=x_0}^{X} \sum_{t=0}^{T-1} D_{i,x,t}}{15\,(X-x_0+1)\,T} \ , \tag{SE19}$$

where *i* ranges over all subjects from the group under consideration, and the 15 arises because each group has 15 subjects.

To obtain R_i^2 for the training exposures, we made $x_0 = 0$ and $X = 3$ (because we numbered training exposures from 0 to 3) and $T = 5$ (because each training exposure had 5 fear ratings). To obtain R_i^2 for the test exposure, we made $x_0 = 4$ and $X = 4$ (because we numbered the test exposure 4) and $T = 6$ (because the test exposure had 6 fear ratings).

FAILURE TO PREDICT TREATMENT OUTCOME WITHOUT THE MODEL

We showed in the main text that the model parameter λ provided useful information to predict treatment outcome, defined as the difference in scores in the Fear of Spiders Questionnaire (FSQ) from pre- to post-treatment (ΔFSQ). Of course, if equally good or even better predictions could be obtained without the model, the model would not be useful for treatment-outcome prediction. We therefore investigated whether we could predict treatment outcome using variables derived directly from the data, without using the model. As for the model-based parameters, we focused on variables derived from only the first two exposures to see if we could predict ultimate treatment outcome early in treatment. To provide a fair comparison to the model-based predictions, we focused on variables that captured similar phenomena to the model parameters. Specifically, we used, for each individual patient, the mean fear decrease in the first two exposures (which, like the model parameter λ, characterizes the level of fear decrease within exposures) and the return of fear between the first and second exposure (which, like the model parameter α, characterizes return of fear).

Exploratory data visualization suggested that ΔFSQ might relate somewhat positively to mean fear decrease after two exposures, but there was no apparent systematic relation between ΔFSQ and return of fear between the first and second exposure (Figure S10). Thus, just as in the model-based predictions we focused on predicting ΔFSQ using only λ through simple linear regression, for the non-model-based predictions we focused on predicting ΔFSQ using only the mean fear decrease in the first two exposures through simple linear regression. Unlike for λ, however, the regression coefficient for the mean fear decrease was not significant ($b = 0.281$, $t = 1.74$, $p = .111$). Moreover, the predicted

R^2 for the regression was negative ($R^2 = .215$, adjusted $R^2 = .144$, predicted $R^2 = –.373$), showing that the mean fear decrease was not a useful predictor.

We also tested, again using a simple linear regression, whether the total fear decrease over the first two exposures (defined as the first fear rating of the first exposure minus the last fear rating of the second exposure) would be useful for treatment-outcome prediction. We selected this variable because it included the effects of both fear decrease within exposures and return of fear between exposures. Again, the coefficient for this variable was not significant ($b = 0.196$, $t = 1.46$, $p = .173$) and the predicted R^2 for the regression was negative ($R^2 = .162$, adjusted $R^2 = .086$, predicted $R^2 = –.277$), showing that this variable was also not useful as a treatment-outcome predictor.

To summarize, trying to predict ΔFSQ using variables derived directly from the data without using the model did not work. Moreover, all values of R^2 for the regression of ΔFSQ on λ ($R^2 = .341$, adjusted $R^2 = .282$, and predicted $R^2 = .128$) were greater than the respective values for the regression of ΔFSQ on mean fear decrease ($R^2 = .215$, adjusted $R^2 = .144$, and predicted $R^2 = –.373$) or on total fear decrease ($R^2 = .162$, adjusted $R^2 = .086$, and predicted $R^2 = –.277$), which shows that the model-derived parameter λ is a better predictor than the non-model based variables mean fear decrease or total fear decrease—or, to use the terminology of dominance analysis, λ dominates mean fear decrease and total fear decrease (16).

SUPPLEMENTAL TABLES

Table S1. Disadvantages of analyses based on classical statistics and how they are overcome by the model-based analyses.

Analyses based on classical statistics are disadvantageous because they:	**Model-based analyses address those disadvantages because they:**
Are unable to capture the detailed patterns of fear change	Use specific mathematical formulations for each component process that underpins fear change, selected from among a set of plausible functions, thereby allowing more precise quantitative descriptions and predictions
Treat each process (e.g., fear decrease and return of fear) in isolation, failing to address how processes interact	Combine the mathematical formulations of each component process into an integrated model
Focus on average effects, so they are unsuited for individual-subject predictions	Provide parameters for each subject, thereby supporting individual-level predictions
May lead to erroneous conclusions if they are not interpreted with care (e.g., concluding that the changes in fear decrease across exposures reflect a parametric change in the underlying process when, as we will show, those changes result from an exponential decay with a constant decay rate)	Support more precise quantitative investigation of the underlying patterns, thereby correcting potentially erroneous conclusions from the classical statistical analyses

Table S2. Quantitative assessment of the model fits and predictions for the training and test exposures, respectively.

		Subject	**Training Exposures**				**Test Exposure**			
		#	**MABS**	**RMSE**	***r***	***R^2***	**MABS**	**RMSE**	***r***	***R^2***
Single-Context Group	Individual Subjects	3	4.28	5.20	0.95	0.77				
		5	3.30	4.21	0.99	0.87				
		9	7.18	7.61	0.96	0.80				
		10	5.28	6.52	0.93	0.88				
		14	5.26	6.34	0.81	0.95				
		22	5.32	6.94	0.92	0.92				
		24	3.68	4.26	0.94	0.97				
		36	5.83	6.48	0.94	0.84				
		38	5.13	6.76	0.93	0.70				
		43	7.44	8.85	0.94	0.59				
		45	4.96	6.12	0.78	0.98				
		47	5.23	6.23	0.97	0.87				
		52	4.22	5.47	0.96	0.96				
		57	3.28	4.17	0.97	0.96				
		63	5.49	6.46	0.90	0.93				
	Group	***Mean***	**5.06**	**6.11**	**0.93**	**0.87**				
		SD	**1.21**	**1.29**	**0.06**	**0.11**				
		Median	**5.23**	**6.34**	**0.94**	**0.88**				
Multiple-Contexts Group	Individual Subjects	2	5.10	6.54	0.85	0.83	4.19	5.44	0.92	0.06
		4	7.85	9.15	0.87	0.93	20.77	23.31	0.98	0.35
		7	6.04	6.87	0.97	0.73	5.96	6.34	0.92	0.83
		8	5.74	6.75	0.95	0.82	6.00	6.42	0.98	0.78
		11	4.08	4.95	0.98	0.97	1.16	1.36	0.92	0.99
		16	11.95	13.37	0.81	0.73	5.44	6.59	0.96	0.78
		19	5.94	7.45	0.96	0.81	2.66	3.36	0.75	0.90
		27	6.76	7.95	0.94	0.90	0.41	0.45	ND[a]	0.99
		30	8.16	10.05	0.72	0.79	2.11	2.80	0.93	0.97
		32	2.77	3.31	0.98	0.98	0.71	0.85	0.94	0.99
		34	9.84	11.32	0.91	0.50	11.30	12.66	0.60	-2.75[b]
		50	12.41	13.53	0.49	0.21	11.50	11.79	0.95	0.64
		51	4.06	4.57	0.98	0.97	11.60	12.52	0.48	-0.19[b]
		53	3.28	4.16	0.63	0.97	8.73	9.15	0.92	0.91
		66	3.63	4.48	0.90	0.97	2.19	2.38	ND[a]	0.98
	Group	***Mean***	**6.51**	**7.63**	**0.86**	**0.81**	**6.32**	**7.03**	**0.87**	**0.48**
		SD	**3.02**	**3.27**	**0.14**	**0.21**	**5.61**	**6.12**	**0.16**	**0.96**
		Median	**5.94**	**6.87**	**0.91**	**0.83**	**5.44**	**6.34**	**0.92**	**0.83**

Note. subject #: subject number; MABS: mean of the absolute values of the residuals; RMSE: root mean squared error; *r*: Pearson correlation coefficient. In nonlinear models, R^2 is not equal to the square of *r* (17), so the two metrics are not redundant. We did not assess accuracy of the predictions for the test exposure for the single-context group because we did not expect the model for that group to generalize well to the test exposure (see text).

[a]The value of *r* is not defined (ND) for subjects #27 and #66 in the test exposure because these subjects had constant fear ratings throughout the test exposure (and therefore 0 variance). The model predictions for those two subjects during the test exposure, however, were very accurate, as can be seen by (a) visual inspection (Figure 5B), (b) the very small MABS and RMSE values, and (c) the very large R^2 values.

[b]With nonlinear models, R^2 can be negative (17); this occurs when the model predictions are worse than those of a null model with constant prediction (see Equation SE18). We found negative values of R^2 only for two subjects during the test exposure: subjects #34 and #51. These subjects exhibited atypical behavior during the test exposure, with increases in fear mid-way through the exposure (Figure 5B).

SUPPLEMENTAL FIGURES

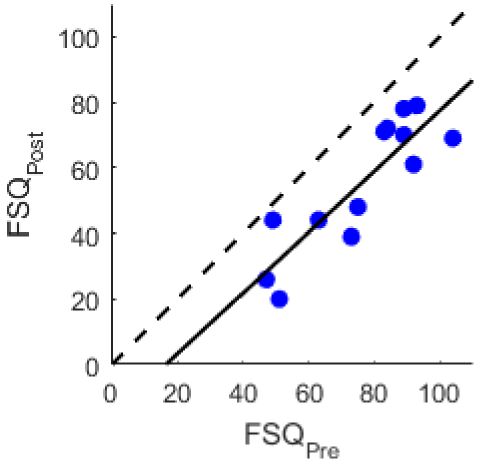

Figure S1. Relation between pre- and post-treatment scores in the Fear of Spiders Questionnaire (FSQ_{Pre} and FSQ_{Post}, respectively) for the subjects in the multiple-contexts group (*n* = 13; although there were 15 subjects in that group, 2 subjects did not have FSQ_{Post} scores). The dashed line has a slope of 1 and an intercept of 0; the fact that all points fall below this line shows that all subjects had lower FSQ_{Post} than FSQ_{Pre} scores. The solid line shows a linear fit.

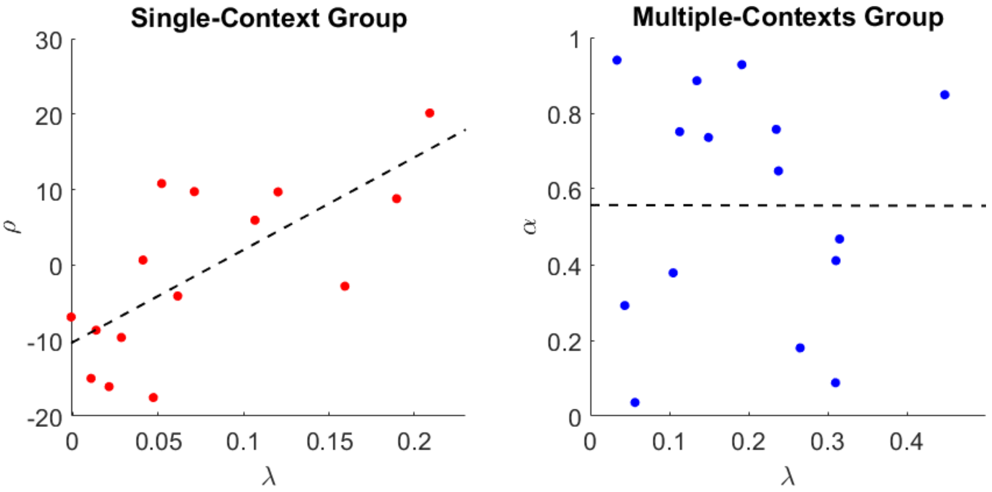

Figure S2. Scatterplots of the parameter estimates for the single-context (left; red) and multiple-contexts (right; blue) groups. The dashed lines represent linear fits.

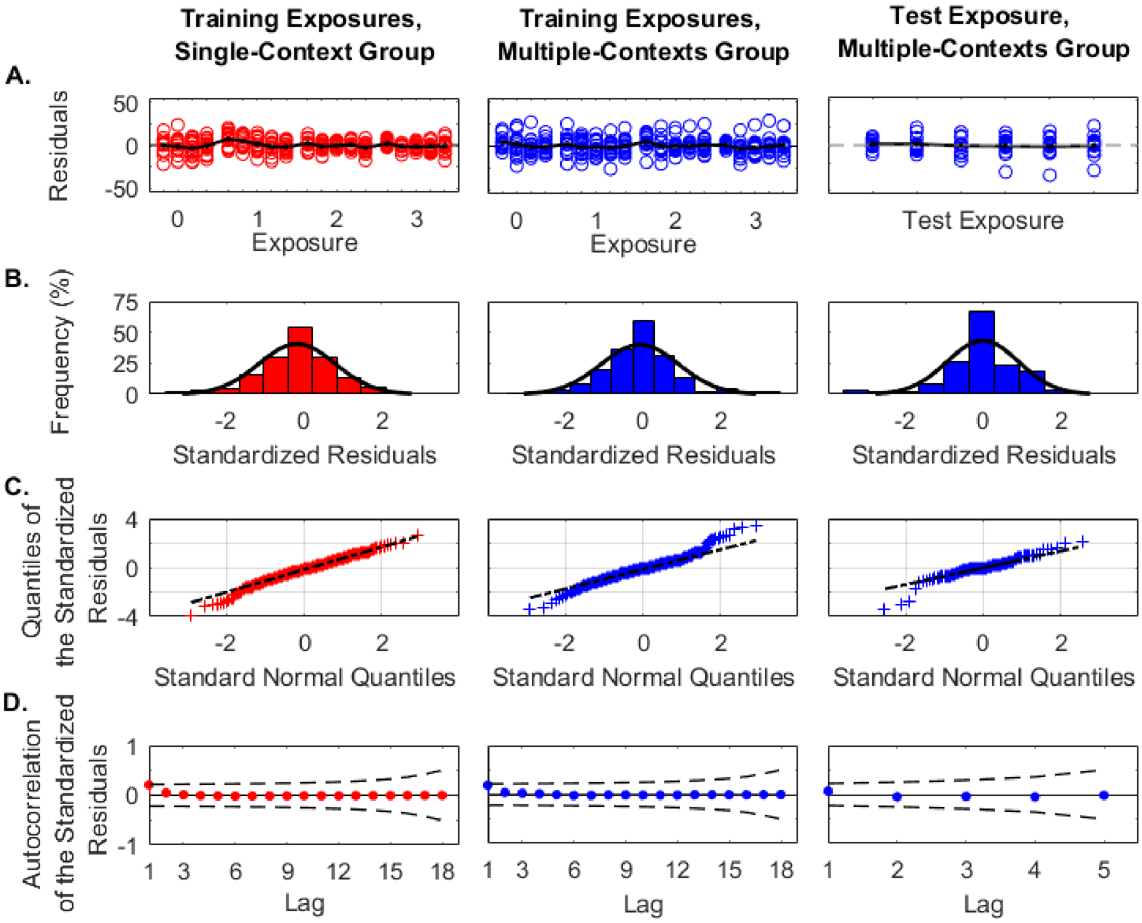

Figure S3. Residuals for the model fits to the fear ratings for the single-context group (left column) and the multiple-contexts group (middle column) during the training exposures and for the model predictions for the multiple-contexts group during the test exposure (right column). The residuals for the first time point of the first training exposure were excluded from all residual plots because, for that specific time point, the model uses the observed data as its estimate (Equations SE2 and SE15; Figure 3), so those residuals are, by definition, 0. **A.** Residuals for each time point of each exposure. Circles indicate residuals; the black line indicates the mean of the residuals at each time point in each exposure. The observed values show no systematic deviations from the model predictions: the mean of the residuals is approximately 0 at every time point, and the distribution of residuals appears largely symmetric around 0 with relatively constant variance. Moreover, residuals that fall farther on the tails of the distribution of residuals generally are attributable to single subjects with atypical behavior. For example, the four large positive residuals in the last four time points of training exposure 3 in the multiple-contexts group all come from subject #50, who showed an atypical large increase in fear during that exposure relative to the preceding exposures (Figure 5B). Similarly, the large negative residuals in the last three time points of the test exposure all come from subject #4, who showed much steeper fear decrease during the test exposure than during the

training exposures, leading the model to predict substantially higher fear than the subject exhibited during the last three time points of the test exposure (Figure 5B). **B.** Histogram of the standardized residuals (bars), with the fitted normal probability density function (black lines). Standardized residuals are residuals normalized to have a *SD* of 1, so a good model should have standardized residuals concentrated around 0, with a symmetric histogram well fit by a normal probability density function, and with most values falling between –2 and +2 (18). In all cases, the standardized residuals for our model fulfill all these characteristics. The only standardized residuals that fall outside the range from –2 to +2 correspond to the outlier residuals from subject #4 during the test exposure. **C.** Normal probability plot of the standardized residuals. The standardized residuals largely fall on a straight line, which further demonstrates that they are well captured by a normal distribution. For the training exposures, there is a hint of slightly heavy tails, especially for the multiple-contexts group, but these are not pronounced (see also panel B). For the test exposure, the three points that deviate from the straight line at the bottom left correspond to the outlier residuals from subject #4. **D.** Autocorrelation of the standardized residuals (dots) with critical limits for a two-sided test of the autocorrelations using $\alpha = 0.05$ (dashed black lines). Nearly all autocorrelations are well inside the critical limits, which demonstrates that generally there is no autocorrelation in the residuals. The only exception concerns the autocorrelations at lag 1 during the training exposures for each group, which are very close to the critical limits; however, even these autocorrelations do not reach statistical significance (single-context group: $\chi^2 = 2.88$, $df = 1$, $p = .090$; multiple-contexts group: $\chi^2 = 2.26$, $df = 1$, $p = .133$). Of course, failure to reject the null hypothesis should not be construed as compelling evidence for that hypothesis, so it is conceivable that the autocorrelations at lag 1 are real. Even if that were the case, however, it would not be indicative of a limitation with the model: departures from model predictions due to noise could well linger for more than a minute—after all, fear is not an instantaneous process—and this would explain the positive correlations at lag 1. Importantly, those correlations would still be induced by noise, so they should not be captured in the model itself (only possibly in the noise model). What would be concerning from the perspective of the model would be if there were other, non-neighbouring autocorrelations suggestive of some systematic failure of the model, which clearly was not the case.

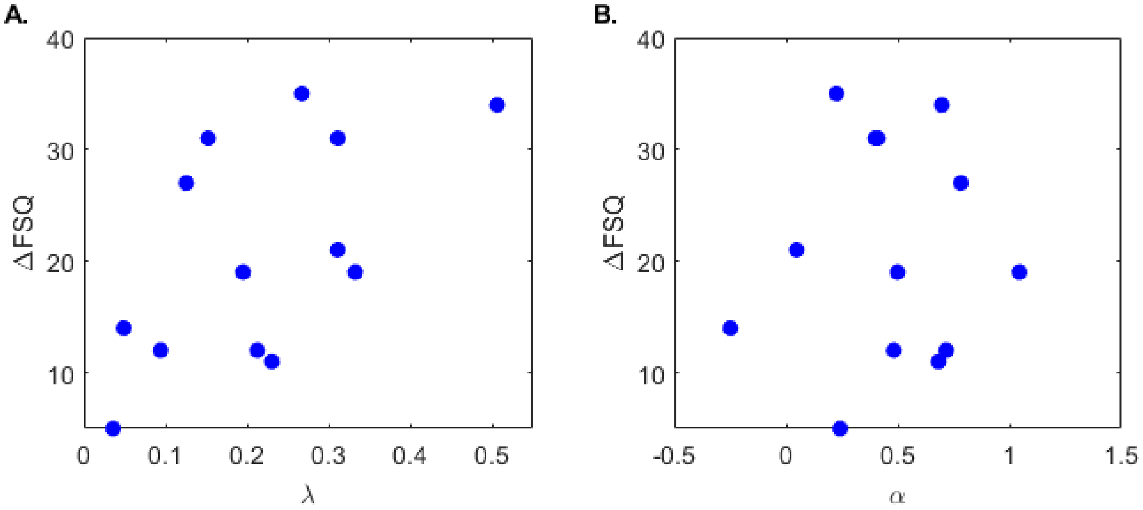

Figure S4. Scatterplots of the change in scores in the Fear of Spiders Questionnaire (FSQ) from pre- to post-treatment (ΔFSQ) versus model parameters estimated using only data from the first two exposures. **A.** Scatterplot of ΔFSQ versus λ. The scatterplot suggests a positive relation between ΔFSQ and λ. **B.** Scatterplot of ΔFSQ versus α. There is no apparent relation between ΔFSQ and α.

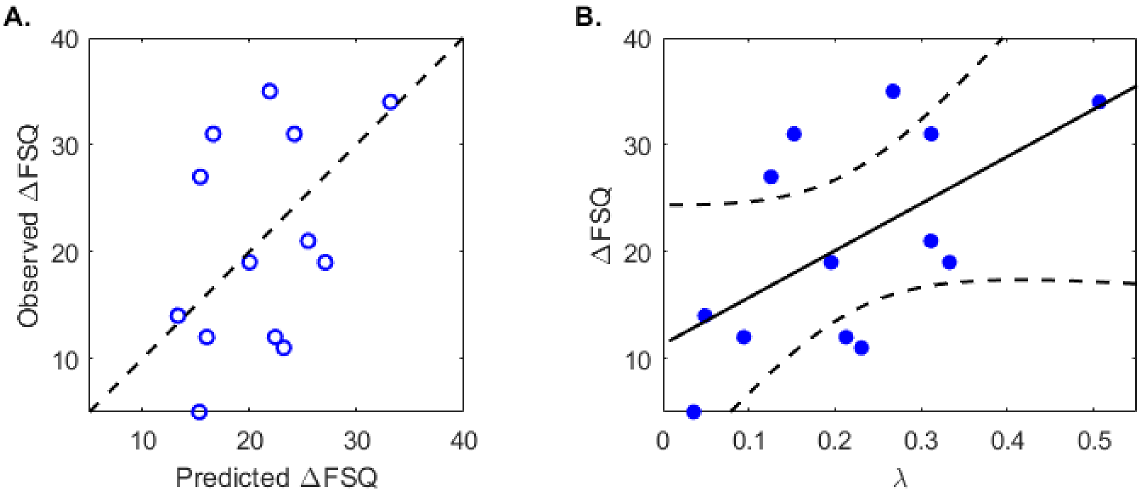

Figure S5. Predictions of the change in scores in the Fear of Spiders Questionnaire (FSQ) from pre- to post-treatment (ΔFSQ) using a simple linear regression with the model parameter λ, estimated using only data from the first two exposures, as the predictor. **A.** Scatterplot of the observed versus predicted ΔFSQ. The predicted ΔFSQ was obtained using leave-one-out cross-validation. The dashed line represents the identity line. **B.** Scatterplot of the observed ΔFSQ versus λ. The solid line represents the regression line; the dashed lines represent 95% simultaneous prediction intervals.

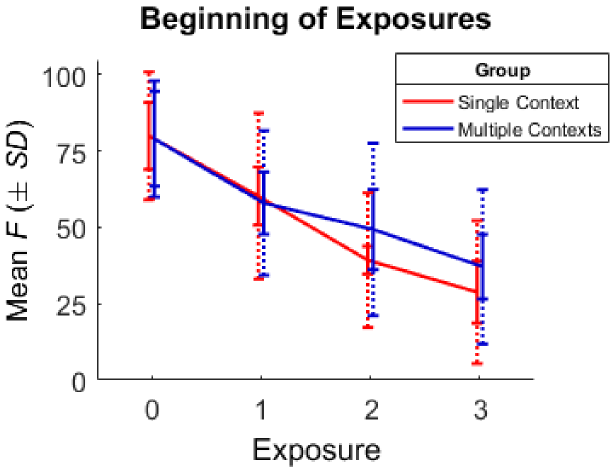

Figure S6. Mean (± *SD*) fear ratings (*F*) at the beginning of each exposure for the single-context (n = 15) and multiple-contexts (n = 15) groups. Dashed error bars represent *SD*s; solid error bars represent *SD*s with the variability related to between-subject differences removed [by mean-normalizing each subject's scores and correcting for the bias introduced by that normalization (19)]. The initial fear ratings decrease across exposures, showing a beneficial, cumulative effect of treatment; they also seem possibly slightly larger for the multiple-contexts than for the single-context group in later exposures (2 and 3), presumably because the multiple-contexts group but not the single-context group has return of fear (Figure 2B). We analyzed these ratings using a mixed ANOVA with Exposure (0, 1, 2, 3) and Group (single context, multiple contexts) as within- and between-subject factors, respectively. We applied Greenhouse-Geisser correction due to violation of the sphericity assumption (1). The decrease in initial fear ratings across exposures was significant (main effect of Exposure: F = 89.27, df = 1.99, 55.97, p < .001, η^2 = .761), exhibiting a significant linear trend both for the two groups combined (linear contrast: F = 151.39, df = 1, 28, p < .001, η^2 = .844) and for each group separately (linear contrast for the single-context group: F = 163.99, df = 1, 14, p < .001, η^2 = .921; linear contrast for the multiple-contexts group: F = 40.58, df = 1, 14, p < .001, η^2= .743). Neither the main effect of Group (F = 0.24, df = 1, 28, p = .631, η^2 = .008) nor the interaction of Group × Exposure (F = 2.20, df = 1.99, 55.97, p = .120, η^2 = .073) reached statistical significance. Still, we also explicitly tested if the higher initial fear ratings observed in the graph for the multiple-contexts group relative to the single-context group in exposures 2 and 3 were significant; that contrast (Group: [–1 1]; Exposure: [0 0 1/2 1/2]) was not statistically significant (F = 1.06, df = 1, 28, p = .311, η^2 = .037, mean difference = 9.20, 95% CI [–9.07, 27.47]). Thus, we cannot conclude that the seemingly slightly larger initial fear ratings in the multiple-contexts group relative to the single-context group in later exposures (2 and 3) indicate a real effect—although we suspect that they do, because of the differences in return of fear between the groups, so we interpret the lack of significant evidence for that effect as probably reflective of insufficient power rather than of absence of the effect.

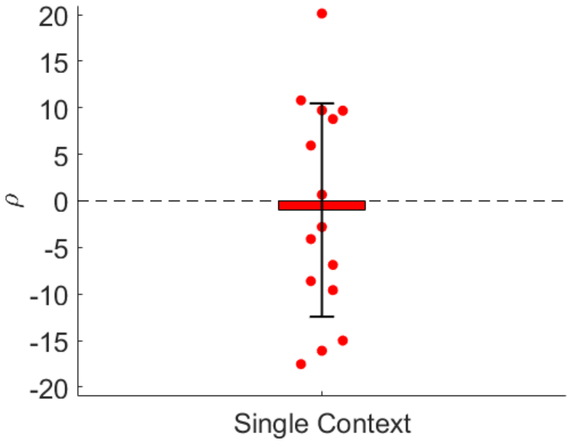

Figure S7. Values of the parameter (ρ) that characterizes the constant return of fear in the model-based analyses of subjects in the single-context group. The values of ρ for each individual subject are shown in red circles (jittered horizontally to avoid overlap, thereby facilitating visualization); the mean (±*SD*) is shown by the red bar. The mean of ρ is very close to 0 (mean = –0.99: less than 1 point in a scale that ranges from 0 to 100), but some subjects have meaningful positive and others have meaningful negative values of ρ. For example, subjects 5, 45, and 24 show consistently negative, null, and positive return of fear, respectively (see Figure 5). The psychological mechanisms underlying negative values of ρ, which reflect a decrease in fear during the exposure interval, are unclear. We suspect that this effect occurs because of the very short exposure interval (2 min) and the fact that all exposures in the single-context group had the same stimulus. A speculative possibility is that subjects who showed this decrease in fear during the exposure interval might have continued to represent the spider internally, thereby continuing fear decrease during the interval. An alternative, or additional, possibility is that this behavior reflects a perceived demand characteristic whereby subjects believe that they are expected to report decreasing fear throughout the exposure treatment.

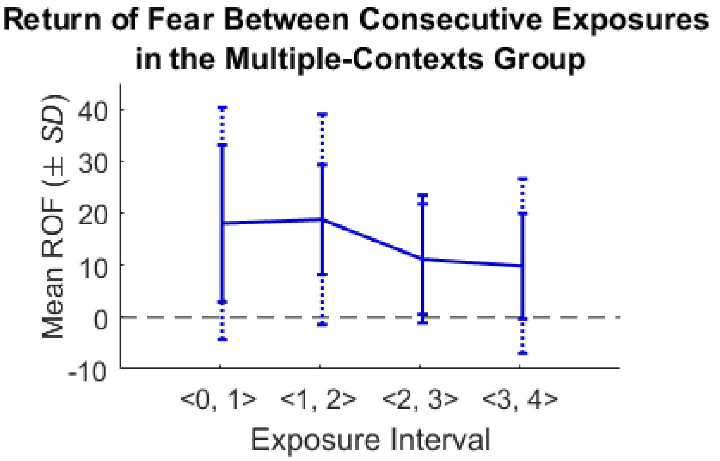

Figure S8. Mean (±*SD*) return of fear (ROF) between consecutive exposures for the multiple-contexts ($n = 15$) group, including ROF from the last training exposure (exposure 3) to the test exposure (exposure 4). Dashed error bars represent *SD*s; solid error bars represent *SD*s with the variability related to between-subject differences removed [by mean-normalizing each subject's scores and correcting for the bias introduced by that normalization (19)]. ROF in the interval between exposures $x - 1$ and x was calculated as fear at the beginning of exposure x minus fear at the end of exposure $x - 1$. We represent the interval between exposures $x - 1$ and x by $<x - 1, x>$. The plot suggests that ROF may decrease across exposure intervals. A one-way repeated measures ANOVA with Exposure Interval (<0, 1>, <1, 2>, <2, 3>, <3, 4>) as the within-subjects factor did not provide statistically significant evidence that ROF varied across exposure intervals ($F = 2.15$, df =3, 42, $p = .109$, $\eta^2 = .133$). A linear contrast, however, was almost significant ($F = 4.161$, $df = 1, 14$, $p = .061$, $\eta^2 = .229$), and a contrast between the first two vs. the last two exposure intervals was statistically significant ([1/2 1/2 –1/2 –1/2]: $F = 5.39$, $df = 1, 14$, $p = .036$, $\eta^2 = .278$, mean difference = 7.90, 95% CI [0.60, 15.20]). The results of these contrasts therefore provide some evidence for the hypothesis that ROF decreased across exposure intervals, although such evidence must, of course, be considered tentative because the omnibus test for the repeated-measures ANOVA did not reach statistical significance.

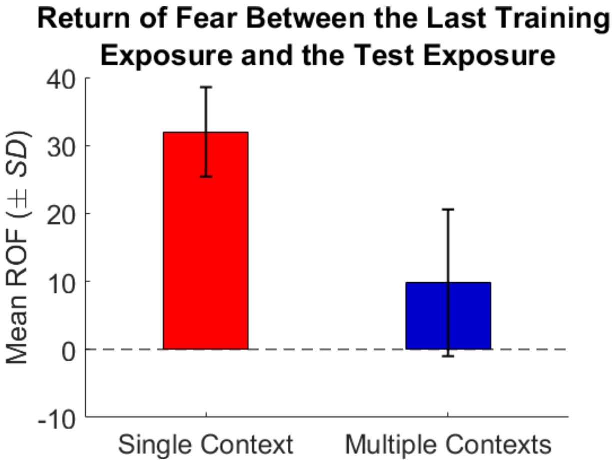

Figure S9. Mean return of fear (ROF) (± *SD*) between the last training exposure and the test exposure for the single-context ($n = 15$) and multiple-contexts ($n = 15$) groups. Although ROF was significantly greater than 0 in both groups (one-sample *t*-test for the single-context group: $t = 6.56$, $df = 14$, $p <. 001$, mean = 32.00, 95% CI [21.55, 42.46]; one-sample *t*-test for the multiple-contexts group: $t = 2.26$, $df = 14$, $p = .040$, mean = 9.80, 95% CI [0.50, 19.10]), it was significantly greater in the single-context group than in the multiple-contexts group (independent-samples *t*-test: $t = 2.58$, $df = 28$, $p = .015$, mean difference single-context group minus multiple-contexts group: 21.53, 95% CI [4.43, 38.64]).

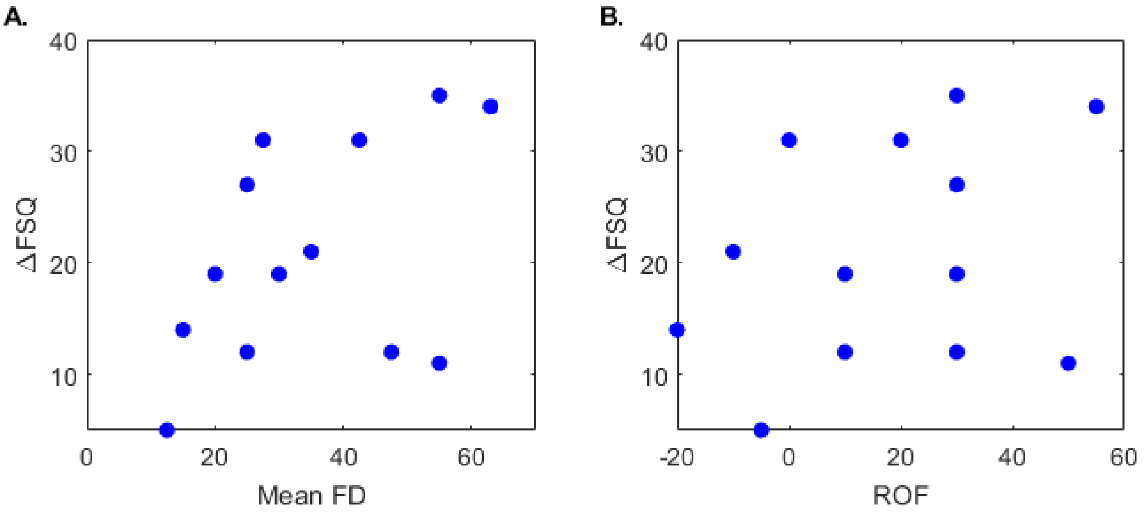

Figure S10. Scatterplots of the change in scores in the Fear of Spiders Questionnaire (FSQ) from pre- to post-treatment (ΔFSQ) versus variables obtained directly from the data of the first two exposures, without using the model. **A.** Scatterplot of ΔFSQ versus the mean fear decrease (FD) in the first two exposures. The scatterplot might be suggestive of a positive relation between ΔFSQ and mean FD. **B.** Scatterplot of ΔFSQ versus the return of fear (ROF) from the first to the second exposure. There is no apparent relation between ΔFSQ and ROF.

SUPPLEMENTAL REFERENCES

1. Park E, Cho M, Ki C-S (2009): Correct use of repeated measures analysis of variance. Korean J Lab Med 29:1–9.
2. Bouton ME (2004): Context and behavioral processes in extinction. Learn Mem 11:485–94.
3. McSweeney FK, Swindell S (2002): Common processes may contribute to extinction and habituation. J Gen Psychol 129:364–400.
4. Balooch SB, Neumann DL, Boschen MJ (2012): Extinction treatment in multiple contexts attenuates ABC renewal in humans. Behav Res Ther 50:604–609.
5. Andreatta M, Leombruni E, Glotzbach-Schoon E, Pauli P, Mühlberger A (2015): Generalization of contextual fear in humans. Behav Ther 46:583–596.
6. Thomas BL, Vurbic D, Novak C (2009): Extensive extinction in multiple contexts eliminates the renewal of conditioned fear in rats. Learn Motiv 40:147–159.
7. Anderson JR (2015): Cognitive Psychology and Its Implications, 8th ed. New York, NY: Worth Publishers.
8. Newell A, Rosenbloom PS (1981): Mechanisms of skill acquisition and the law of practice. In: Anderson JR, editor. Cognitive Skills and Their Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates, 1–55.
9. Rankin CH, Abrams T, Barry RJ, Bhatnagar S, Clayton DF, Colombo J, et al. (2009): Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. Neurobiol Learn Mem 92:135–138.
10. Rubin DC, Wenzel AE, Rubin DC, Psychology DOE, Hinton S, Machado O, et al. (1996): One hundred years of forgetting: A quantitative description of retention. Psychol Rev 734–760.
11. Staddon JER, Higa JJ (1999): Time and memory: Towards a pacemaker-free theory of interval timing. J Exp Anal Behav 71:215–51.
12. Pavlov IP (1927): Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. London: Oxford University Press.
13. van der Schaft A, Schumacher H (2000): An Introduction to Hybrid Dynamical Systems. Lecture Notes in Control and Information Sciences. (Vol. 251), London: Springer London. doi: 10.1007/BFb0109998.
14. Daunizeau J, Adam V, Rigoux L (2014): VBA: A probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comput Biol 10:e1003441.
15. Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014): Bayesian model selection for group studies — Revisited. Neuroimage 84:971–985.
16. Budescu D V (1993): Dominance analysis: A new approach to the problem of relative

importance of predictors in multiple regression. Psychol Bull 114:542–551.

17. Motulsky H, Christopoulos A (2004): Fitting Models to Biological Data using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting. Oxford University Press.
18. Mould DR, Upton RN (2013): Basic concepts in population modeling, simulation, and model-based drug development. Part 2: Introduction to pharmacokinetic modeling methods. CPT Pharmacometrics Syst Pharmacol 2:e38.
19. Morey RD (2008): Confidence intervals from normalized data: a correction to Cousineau (2005). Tutor Quant Methods Psychol 4:61–64.